

# TRANSPORTATION RESEARCH RECORD

---

Journal of the Transportation Research Board, No. 2261

Railways  
2011

*Including*

2011 Thomas B. Deen  
Distinguished Lecture

**TRANSPORTATION RESEARCH RECORDS**, which are published throughout the year, consist of collections of papers on specific transportation modes and subject areas. Each Record is classified according to the subscriber category or categories covered by the papers published in that volume. The views expressed in papers published in the Transportation Research Record series are those of the authors and do not necessarily reflect the views of the peer review committee(s), the Transportation Research Board, the National Research Council, or the sponsors of TRB activities. The Transportation Research Board does not endorse products or manufacturers; trade and manufacturers' names may appear in a Record paper only if they are considered essential.

**PEER REVIEW OF PAPERS:** All papers published in the Transportation Research Record series have been reviewed and accepted for publication through the Transportation Research Board's peer review process established according to procedures approved by the Governing Board of the National Research Council. Papers are refereed by the TRB standing committees identified on page ii of each Record. Reviewers are selected from among committee members and other outside experts. The Transportation Research Board requires a minimum of three reviews; a decision is based on reviewer comments and resultant author revision. For details about the peer review process, see the information on the inside back cover.

**THE TRANSPORTATION RESEARCH RECORD PUBLICATION BOARD**, comprising a cross section of transportation disciplines and with equal representation from the academic and practitioner communities, assures that the quality of the *Transportation Research Record: Journal of the Transportation Research Board* and the TRB paper peer review process is consistent with that of a peer-reviewed scientific journal. Members from the academic community make decisions on granting tenure; all academic and practitioner members have authored papers published in peer-reviewed journals and all have participated in the TRB peer review process.

**TRANSPORTATION RESEARCH RECORD PAPERS ONLINE:** The TRR Journal Online website provides electronic access to the full text of papers that have been published in the Transportation Research Record series since 1996. The site is updated as new Record papers become available. To search abstracts and to find subscription and pricing information, go to [www.TRB.org/TRROnline](http://www.TRB.org/TRROnline).

**TRANSPORTATION RESEARCH BOARD PUBLICATIONS** may be ordered directly from the TRB Business Office, through the Internet at [www.TRB.org](http://www.TRB.org), or by annual subscription through organizational or individual affiliation with TRB. Affiliates and library subscribers are eligible for substantial discounts. For further information, contact the Transportation Research Board Business Office, 500 Fifth Street, NW, Washington, DC 20001 (telephone 202-334-3213; fax 202-334-2519; or e-mail [TRBsales@nas.edu](mailto:TRBsales@nas.edu)).

## TRANSPORTATION RESEARCH BOARD 2011 EXECUTIVE COMMITTEE\*

**Chair:** Neil J. Pedersen, Consultant, Silver Spring, Maryland

**Vice Chair:** Sandra Rosenbloom, Professor of Planning, University of Arizona, Tucson

**Executive Director:** Robert E. Skinner, Jr., Transportation Research Board

**J. Barry Barker**, Executive Director, Transit Authority of River City, Louisville, Kentucky  
**Deborah H. Butler**, Executive Vice President, Planning, and CIO, Norfolk Southern Corporation, Norfolk, Virginia  
**William A. V. Clark**, Professor, Department of Geography, University of California, Los Angeles  
**Eugene A. Conti, Jr.**, Secretary of Transportation, North Carolina Department of Transportation, Raleigh  
**James M. Crites**, Executive Vice President of Operations, Dallas-Fort Worth International Airport, Texas  
**Paula J. Hammond**, Secretary, Washington State Department of Transportation, Olympia  
**Michael W. Hancock**, Secretary, Kentucky Transportation Cabinet, Frankfort  
**Adib K. Kanafani**, Professor of the Graduate School, University of California, Berkeley (Past Chair, 2009)  
**Michael P. Lewis**, Director, Rhode Island Department of Transportation, Providence  
**Susan Martinovich**, Director, Nevada Department of Transportation, Carson City  
**Joan McDonald**, Commissioner, New York State Department of Transportation, Albany  
**Michael R. Morris**, Director of Transportation, North Central Texas Council of Governments, Arlington (Past Chair, 2010)  
**Tracy L. Rosser**, Vice President, Regional General Manager, Wal-Mart Stores, Inc., Mandeville, Louisiana  
**Steven T. Scalzo**, Chief Operating Officer, Marine Resources Group, Seattle, Washington  
**Henry G. (Gerry) Schwartz, Jr.**, Chairman (retired), Jacobs/Sverdrup Civil, Inc., St. Louis, Missouri  
**Beverly A. Scott**, General Manager and Chief Executive Officer, Metropolitan Atlanta Rapid Transit Authority, Atlanta, Georgia  
**David Seltzer**, Principal, Mercator Advisors LLC, Philadelphia, Pennsylvania  
**Lawrence A. Selzer**, President and CEO, The Conservation Fund, Arlington, Virginia  
**Kumares C. Sinha**, Olson Distinguished Professor of Civil Engineering, Purdue University, West Lafayette, Indiana  
**Thomas K. Sorel**, Commissioner, Minnesota Department of Transportation, St. Paul  
**Daniel Sperling**, Professor of Civil Engineering and Environmental Science and Policy; Director, Institute of Transportation Studies; and Interim Director, Energy Efficiency Center, University of California, Davis  
**Kirk T. Steudle**, Director, Michigan Department of Transportation, Lansing  
**Douglas W. Stotlar**, President and Chief Executive Officer, Con-Way, Inc., Ann Arbor, Michigan  
**C. Michael Walton**, Ernest H. Cockrell Centennial Chair in Engineering, University of Texas, Austin (Past Chair, 1991)

**Rebecca M. Brewster**, President and COO, American Transportation Research Institute, Smyrna, Georgia (ex officio)  
**Anne S. Ferro**, Administrator, Federal Motor Carrier Safety Administration, U.S. Department of Transportation (ex officio)  
**LeRoy Gishi**, Chief, Division of Transportation, Bureau of Indian Affairs, U.S. Department of the Interior, Washington, D.C. (ex officio)  
**John T. Gray**, Senior Vice President, Policy and Economics, Association of American Railroads, Washington, D.C. (ex officio)  
**John C. Horsley**, Executive Director, American Association of State Highway and Transportation Officials, Washington, D.C. (ex officio)  
**Michael P. Huerta**, Acting Administrator, Federal Aviation Administration, U.S. Department of Transportation (ex officio)  
**David T. Matsuda**, Deputy Administrator, Maritime Administration, U.S. Department of Transportation (ex officio)  
**Michael P. Melaniphy**, President, American Public Transportation Association, Washington, D.C. (ex officio)  
**Victor M. Mendez**, Administrator, Federal Highway Administration, U.S. Department of Transportation (ex officio)  
**Tara O'Toole**, Under Secretary for Science and Technology, U.S. Department of Homeland Security (ex officio)  
**Robert J. Papp** (Adm., U.S. Coast Guard), Commandant, U.S. Coast Guard, U.S. Department of Homeland Security (ex officio)  
**Cynthia L. Quarterman**, Administrator, Pipeline and Hazardous Materials Safety Administration, U.S. Department of Transportation (ex officio)  
**Peter M. Rogoff**, Administrator, Federal Transit Administration, U.S. Department of Transportation (ex officio)  
**David L. Strickland**, Administrator, National Highway Traffic Safety Administration, U.S. Department of Transportation (ex officio)  
**Joseph C. Szabo**, Administrator, Federal Railroad Administration, U.S. Department of Transportation (ex officio)  
**Polly Trottenberg**, Assistant Secretary for Transportation Policy, U.S. Department of Transportation (ex officio)  
**Robert L. Van Antwerp** (Lt. General, U.S. Army), Chief of Engineers and Commanding General, U.S. Army Corps of Engineers, Washington, D.C. (ex officio)  
**Barry R. Wallerstein**, Executive Officer, South Coast Air Quality Management District, Diamond Bar, California (ex officio)  
**Gregory D. Winfree**, Acting Administrator, Research and Innovative Technology Administration, U.S. Department of Transportation (ex officio)

\* Membership as of December 2011.

# TRANSPORTATION RESEARCH RECORD

---

Journal of the Transportation Research Board, No. 2261

## Railways 2011

*Including*

2011 Thomas B. Deen  
Distinguished Lecture

A Peer-Reviewed Publication

---

TRANSPORTATION RESEARCH BOARD  
*OF THE NATIONAL ACADEMIES*

Washington, D.C.

2011

[www.TRB.org](http://www.TRB.org)

## **Transportation Research Record 2261**

ISSN 0361-1981

ISBN 978-0-309-22296-9

### **Subscriber Categories**

Railroads; passenger transportation; freight transportation

Printed in the United States of America

## **Peer Review of Transportation Research Record 2261**

### **POLICY AND ORGANIZATION GROUP**

Cynthia Jane Burbank, Parsons Brinckerhoff, Inc. (Chair)

### **Transportation Policy Section**

Eric N. Schreffler, Eric Schreffler, Transportation Consultant (Chair)

### **Transportation Economics Committee**

Mark W. Burris, Texas A&M University (Chair), Brian Baird Alstadt, Eric Asa, Rabinder K. Bains, Daniel I. Brod, Sujit Das, Matthieu de Lapparent, William Denning, Paula E. Dowell, Ralph C. Erickson, Eric C. Gabler, Konstantina Gkritza, Daniel Joseph Graham, David T. Hartgen, Marwan F. Madi, John S. Merriss, Jung Eun Oh, Jonathan Richard Peters, Jolanda P. Prozzi, Peter C. Rafferty, Rimona Rafiah, David G. Ripplinger, Anthony M. Rufolo, Mishra Sabyasachee, Elena Safirova, Georgina Santos, Darren P. Timothy, Valentin G. Vulov, Christopher R. Williges, Rhonda K. Young, Lei Zhang

### **RAIL GROUP**

Daniel L. Roth, Ernst & Young Orenda Corporate Finance Inc. (Chair), Dwight W. Clark, David B. Clarke, David D. Davis, John T. Gray II, John Kevin Kesler, Yung-Cheng Lai, Kevin Neels, Anthony D. Perl, Stephen M. Popkin, Stephen J. Reinach, Patrick B. Simmons, Bruce R. Smith, John M. Tunna, David C. Wilcock

### **Intercity Passenger Rail Committee**

Anthony D. Perl, Simon Fraser University (Chair), Camille Tsao, HNTB Corporation (Secretary), Rohit T. Aggarwala, Daniel Brand, Ross B. Capon, Rod J. Diridon, Moshe Givoni, Chelsea L. Gleis, Dharm Guruswamy, George Haikalis, Olivier Klein, Alexander Lu, Jason J. Maga, Deborah Wood Matherly, Ronald A. Mauri, Matthew James Melzer, Curtis A. Morgan, Andrew B. Nash, Howard R. Permut, Eric C. Peterson, Michael Angel Rodriguez, Daniel L. Roth, Allan Rutter, Michael Huntly Schabas, David P. Simpson, Reed H. Tanger, John C. Tone, Eric S. Tyrer II, Randall E. Wade, Mark C. Walbrun

### **Passenger Rail Equipment and Systems Integration Committee**

Patrick B. Simmons, North Carolina Department of Transportation (Chair), Eloy E. Martinez, LTK Engineering Services, Inc. (Secretary), John G. Bell, Alan J. Bing, Robert M. Dorer, Nora C. Friend, Andrew J. Galloway, John A. Harrison, Joseph Lee Hutchins, Jr., Larry David Kelterborn, Andrés López-Pita, Nobuyuki Matsumoto, David O. Nelson, Philip Olekszyk, R. Scott Phelan, Charles A. Poltenson, Sr., Sunduck Daniel Suh, Michael J. Trosino, David A. Valenstein, Robert B. Watson

### **Railroad Operating Technologies Committee**

Yung-Cheng Lai, National Taiwan University (Chair), Adrian D. Hellman, Volpe National Transportation Systems Center (Secretary), Christopher P. L. Barkan, Virginia M. Beck, Stephen Juan Bruno, Olga K. Cataldi, David B. Clarke, Richard U. Cogswell, Llewellyn C. Davis, Timothy J. DePaepe, Mark H. Dingler, Mark W. Hemphill, Ecton K. Holt, Michael Iden, Edwin R. Kraft, Coleman Lawrence, Robert H. Leilich, Marco M. Luethi, Larry R. Milhon, Jena C. Montgomery, Gordon B. Mott, Joern Pacht, Paul H. Reistrup, Mark K. Ricci, Ismail Sahin, Victor Simuoli, Michael E. Smith, Mark P. Stehly, James A. Stem, Jr.

### **Local and Regional Rail Freight Transport Committee**

David B. Clarke, University of Tennessee, Knoxville (Chair), Tim Andrews, C. Howard Capito, Diane Davidson, John M. Gibson, Jr., John T. Gray II, George Avery Grimes, Donald B. Ludlow, R. Leo Penne, Charles A. Poltenson, Sr., Henry Posner III, George Raymond, Randolph R. Resor, Richard Durward Robey, James H. M. Savage, George W. Schafer III, Forrest R. Van Schwartz, David P. Simpson, Anne Strauss-Wieder, Peter F. Swan

### **Railroad Track Structure System Design Committee**

David D. Davis, Transportation Technology Center, Inc. (Chair), Dwight W. Clark, Union Pacific Railroad Company (Secretary), Ernest J.

Barenberg, Timothy R. Bennett, David N. Bilow, Randy L. Bowman, Michael O. Brown, Miodrag Budisa, William G. Byers, Laurence E. Daniels, Riley Edwards, Konstantinos Giannakos, Carlton L. Ho, Henry M. Lees, Jr., Andrés López-Pita, Dale W. Ophardt, Donald Plotkin, Juanjuan Ren, David E. Staplin, Yu-Jiang Zhang

### **Railway Maintenance Committee**

Dwight W. Clark, Union Pacific Railroad Company (Chair), James P. Hyslip, HyGround Engineering, LLC (Secretary), Miodrag Budisa, Gary A. Carr, Thomas F. DeJoseph, Willem Ebersohn, Riley Edwards, Gurnel S. Ghataora, P. J. Hannes Gräbe, Carlton L. Ho, Hai Huang, Stanislav Jovanovic, David L. Kannenberg, Francesco Lanza di Scalea, Eric E. Magel, Valerie Marcolongo, Daniel B. Mesnick, Thomas H. O'Brien, Bruce Pohlot, Christian Mark Roberts, Michael D. Roney, Mario A. Ruel, Conrad J. Ruppert, Jr., Ernest T. Selig, Eric Sherrock, Theodore R. Sussmann, Jr., Vincent R. Terrill, Jackie van der Westhuizen, Timothy R. Wells, Allan M. Zarembski

### **Railroad Operational Safety Committee**

Stephen J. Reinach (Chair), Stephen M. Popkin, Research and Innovative Technology Administration (Vice Chair), Gina M. Melnik, Research and Innovative Technology Administration (Secretary), Michael Copen, Timothy J. DePaepe, Lawrence B. Fleischer, Charles J. Fraley, Frederick C. Gamst, Judith B. Gertler, Kelly Haley, Dennis W. Holland, Rick Inclima, Scott Kaye, Vijay K. Kohli, Jennifer E. Lincoln, David H. Mangold, Ann M. Mills, Jeffrey Franklin Moller, Joern Pacht, Thomas A. Pontolillo, Thomas G. Raslear, Mark K. Ricci, James A. Stem, Jr., Thomas Streicher, Barry L. Wells

### **FREIGHT SYSTEMS GROUP**

Peter F. Swan, Pennsylvania State University, Harrisburg (Chair)

### **Freight Transportation Economics and Regulation Committee**

Kevin Neels, The Brattle Group (Chair), Catherine L. Taylor, Volpe National Transportation Systems Center (Secretary), Michael H. Belzer, Michael A. Benouaich, Stephen V. Burks, David B. Clarke, Deborah M. Freund, Konstantina Gkritza, John T. Gray II, Mark A. Hornung, Mario Iacobacci, Joshua L. Schank, John H. Siebert, Michael E. Smith, Joseph A. Swanson, Louis S. Thompson, Steven J. Tripp, Karen E. White, Peter Wolff

### **Transportation of Hazardous Materials Committee**

Mark Lepofsky, Visual Risk Technologies, Inc. (Chair), Richard C. Bornhorst, U.S. Coast Guard (Secretary), Mark D. Abkowitz, Laura L. Barhydt, Paul M. Bomgardner, Christopher J. Bonanti, Cheryl A. Burke, Timothy Butters, Edward R. Chapman, John L. Conley, John V. Currie, Philip Joseph Daum, Ronald J. Duych, Robert E. Fronczak, Emily K. Goodenough, Scott C. Gorton, Charles H. Hochman, Joseph Gary Lanthrum, Richard Moskowitz, Eftihia G. Nathanail, Steven L. Niswander, Mohd Rapik Saat, John Nmn Salvesen, William S. Schoonover, James O'Neal Simmons, Stephen F. Urschel, Robert D. Waters, Yuanchang Xie

Peer review is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 2010.

### **Transportation Research Board Staff**

Ann R. Purdue, Senior Program Officer, Rail and Freight  
Martine Micozzi, Senior Program Officer, Management, Policy, and International Relations  
Matthew A. Miller, Senior Program Associate  
Richard F. Pain, Senior Program Officer and Transportation Safety Coordinator  
Joanice L. Cole, Senior Program Assistant

### **Publications Office**

Elaine Eldridge, Editor; Julie M. Conroy, Production Editor;  
Deborah Uffelman, Proofreader; A. Regina Reid, Manuscript Preparer  
Ann E. Petty, Managing Editor; Juanita Green, Production Manager;  
Phyllis Barber, Publishing Services Manager; Jennifer J. Weeks, Manuscript Preparation Manager

# THE NATIONAL ACADEMIES

## *Advisers to the Nation on Science, Engineering, and Medicine*

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both the Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation.  
**[www.TRB.org](http://www.TRB.org)**

**[www.national-academies.org](http://www.national-academies.org)**

#### TRANSPORTATION RESEARCH RECORD PUBLICATION BOARD

C. Michael Walton, PhD, PE, Ernest H. Cockrell Centennial Chair  
in Engineering and Professor of Civil Engineering, University of Texas,  
Austin (Cochair)  
Mary Lynn Tischer, PhD, Director, Office of Transportation Policy  
Studies, Federal Highway Administration, Washington, D.C. (Cochair)  
Daniel Brand, SB, SM, PE, Consultant, Lyme, New Hampshire  
Mary R. Brooks, BOT, MBA, PhD, William A. Black Chair of  
Commerce, Dalhousie University, Halifax, Nova Scotia, Canada  
Charles E. Howard, Jr., MCRP, Director, Transportation Planning,  
Puget Sound Regional Council, Seattle, Washington  
Thomas J. Kazmierowski, PEng, Manager, Materials Engineering and  
Research Office, Ontario Ministry of Transportation, Toronto, Canada  
Michael D. Meyer, PhD, PE, Frederick R. Dickerson Professor, School of  
Civil and Environmental Engineering, Georgia Institute of Technology,  
Atlanta  
Sandra Rosenbloom, MA, PhD, Professor of Planning, University of  
Arizona, Tucson  
Kumares C. Sinha, PhD, PE, Olson Distinguished Professor, School of  
Civil Engineering, Purdue University, West Lafayette, Indiana  
L. David Suits, MSCE, Executive Director, North American  
Geosynthetics Society, Albany, New York

#### TRB SPONSORS\*

Transportation Departments of the 50 States and the District  
of Columbia

##### *Federal Government*

U.S. Department of Transportation  
Federal Aviation Administration  
Federal Highway Administration  
Federal Motor Carrier Safety Administration  
Federal Railroad Administration  
Federal Transit Administration  
National Highway Traffic Safety Administration  
Research and Innovative Technology Administration

Bureau of Indian Affairs  
Office of Naval Research, U.S. Navy  
Science and Technology Directorate,  
U.S. Department of Homeland Security  
U.S. Army Corps of Engineers  
U.S. Coast Guard

##### *Nongovernmental Organizations*

American Association of State Highway  
and Transportation Officials  
American Public Transportation Association  
American Transportation Research Institute  
Association of American Railroads  
South Coast Air Quality Management District

---

\*As of December 2011.

# TRANSPORTATION RESEARCH RECORD

Journal of the Transportation Research Board, No. 2261

## Contents

Foreword	ix
<hr/>	
<b>PART 1—2011 THOMAS B. DEEN DISTINGUISHED LECTURE</b>	
James W. McClellan, 2011 Thomas B. Deen Distinguished Lecturer	3
<hr/>	
Railroads and the New Normal: Impact of Lean and Green on the Future of Railroads James W. McClellan	5
<hr/>	
<b>PART 2—RAILWAYS 2011</b>	
Assessment of Wider Economic Impacts of High-Speed Rail for Great Britain Daniel J. Graham and Patricia C. Melo	15
<hr/>	
Economic Impacts of Intercity Passenger Rail Service: Evidence from Passenger Surveys Benjamin R. Sperry and Curtis A. Morgan	25
<hr/>	
Cluster Analysis of Intercity Rail Passengers in Emerging High-Speed Rail Corridor Benjamin R. Sperry, Kristopher D. Ball, and Curtis A. Morgan	31
<hr/>	
High-Speed Railways in Spain: Example of Success? M. Sánchez-Borràs, F. Robusté, and O. Criado	39
<hr/>	
Hierarchical Network Model of Safe High-Speed Rail Operation Yanhui Wang, Xuemei Xiao, Limin Jia, and Chenchen Zhang	49
<hr/>	

<b>Simplified Approach for Assessing Initial Fire Development and Spread in Passenger Rail Vehicles</b> B. J. Meacham, N. A. Dembsey, M. Johann, J. Tubbs, and K. Schebel	<b>57</b>
<b>Geometry of High-Speed Turnouts</b> Miguel R. Bugarín, Alfonso Orro, and Margarita Novales	<b>64</b>
<b>International Variation in Cost-Benefit Analysis of Urban Rail Projects: Impact on Outcomes</b> Evan Gwee, Graham Currie, and John Stanley	<b>73</b>
<b>Influence of Value of Time on Profitability of Railway Projects: Theoretical Formulation and Case Study</b> O. Criado, M. Sánchez-Borràs, and F. Robusté	<b>86</b>
<b>Expanding Alaska-Canada Rail: Jointly Visualizing Revenue Freight, Development Cost, Mineral Commodity Value, and Impact of Carbon Dioxide</b> Colin Brooks, Helen Kourous-Harrigan, Mike Billmire, Paul Metz, D. Eric Keefauver, Robert Shuchman, Richard Dobson, K. Arthur Endsley, and Mark Taylor	<b>95</b>
<b>Midwest Rail Study: Modeled Near-Field Impacts of Emissions of Fine Particulate Matter from Railyard Activities</b> Stephen N. Feinberg, Varun Yadav, Jeremy G. Heiken, and Jay R. Turner	<b>106</b>
<b>Integrated Optimization Model to Manage Risk of Transporting Hazardous Materials on Railroad Networks</b> Yung-Cheng Lai, Athaphon Kawprasert, Chen-Yu Lin, M. Rapik Saat, Chun-Hao Liang, and Christopher P. L. Barkan	<b>115</b>
<b>Train Delay and Economic Impact of In-Service Failures of Railroad Rolling Stock</b> Bryan W. Schlake, Christopher P. L. Barkan, and J. Riley Edwards	<b>124</b>
<b>Basic Tool Kit for Estimation of Intermodal Rail Cost</b> Dan P. K. Seedah, Robert Harrison, and James R. Blaze	<b>134</b>
<b>Noncontact Ultrasonic Guided-Wave System for Rail Inspection: Update on Project at University of California, San Diego</b> Stefano Coccia, Robert Phillips, Ivan Bartoli, Salvatore Salamone, Francesco Lanza di Scalea, Mahmood Fateh, and Gary Carr	<b>143</b>

<b>Installation of Railroad Wayside Defect Detectors: Multiperiod Design</b>	<b>148</b>
Fan Peng, Xiaopeng Li, and Yanfeng Ouyang	
<hr/>	
<b>Track Maintenance of Heavy Haul Railways with Resilient Versus Stiff Fastenings</b>	<b>155</b>
K. Giannakos	
<hr/>	
<b>Biobjective Optimization Model for Maintenance and Renewal Decisions Related to Rail Track Geometry</b>	<b>163</b>
A. Ramos Andrade and P. Fonseca Teixeira	
<hr/>	
<b>Verification of Box Test Model and Calibration of Finite Element Model: Evaluation of Railroad Ballast Performance</b>	<b>171</b>
Kane C. Bennett, Carlton L. Ho, and Hoang Q. Nguyen	
<hr/>	
<b>Analysis of Derailments by Accident Cause: Evaluating Railroad Track Upgrades to Reduce Transportation Risk</b>	<b>178</b>
Xiang Liu, Christopher P. L. Barkan, and M. Rapik Saat	
<hr/>	



# Foreword

The Thomas B. Deen Distinguished Lectureship recognizes the career contributions and achievements of an individual in one of five areas covered by the Transportation Research Board's Technical Activities Division. Honorees are invited to present overviews of their technical areas, including the evolution, the present status, and the prospects for the future. First awarded in 1992, the lectureship was renamed by the TRB Executive Committee in 2003 in honor of the board's eighth executive director, who served with distinction from 1980 to 1994.

James W. McClellan, vice president, Woodside Consulting Group, Virginia Beach, Virginia, was the recipient of the 2011 Distinguished Lectureship. His paper, "Railroads and the New Normal: Impact of Lean and Green on the Future of Railroads," was sponsored by the TRB Technical Activities Council and presented at TRB's 90th Annual Meeting in January 2011. The final, edited paper appears as Part 1 of this volume.

The 2011 series of the *Transportation Research Record: Journal of the Transportation Research Board* consists of approximately 990 papers selected from 3,900 submissions after rigorous peer review. The peer review for each paper published in this volume was coordinated by the committee acknowledged at the end of the text; members of the reviewing committees for the papers in this volume are listed on page ii.

Additional information about the *Transportation Research Record: Journal of the Transportation Research Board* series and the peer review process appears on the inside back cover. TRB appreciates the interest shown by authors in offering their papers, and the board looks forward to future submissions.

**Note:** Many of the photographs, figures, and tables in this volume have been converted from color to grayscale for printing. The electronic files of the papers, posted on the web at [www.TRB.org/TRROnline](http://www.TRB.org/TRROnline), retain the color versions of photographs, figures, and tables as originally submitted for publication.

## Measurement Conversion Factors

To convert from the unit in the first column to the unit in the second column, multiply by the factor in the third column.

<i>Customary Unit</i>	<i>SI Unit</i>	<i>Factor</i>
<b>Length</b>		
inches	millimeters	25.4
inches	centimeters	2.54
feet	meters	0.305
yards	meters	0.914
miles	kilometers	1.61
<b>Area</b>		
square inches	square millimeters	645.1
square feet	square meters	0.093
square yards	square meters	0.836
acres	hectares	0.405
square miles	square kilometers	2.59
<b>Volume</b>		
gallons	liters	3.785
cubic feet	cubic meters	0.028
cubic yards	cubic meters	0.765
<b>Mass</b>		
ounces	grams	28.35
pounds	kilograms	0.454
short tons	megagrams	0.907
<b>Illumination</b>		
footcandles	lux	10.76
footlamberts	candelas per square meter	3.426
<b>Force and Pressure or Stress</b>		
poundforce	newtons	4.45
poundforce per square inch	kilopascals	6.89
<b>Temperature</b>		

To convert Fahrenheit temperature (°F) to Celsius temperature (°C), use the following formula:

$$^{\circ}\text{C} = (^{\circ}\text{F} - 32)/1.8$$

<i>SI Unit</i>	<i>Customary Unit</i>	<i>Factor</i>
<b>Length</b>		
millimeters	inches	0.039
centimeters	inches	0.394
meters	feet	3.281
meters	yards	1.094
kilometers	miles	0.621
<b>Area</b>		
square millimeters	square inches	0.00155
square meters	square feet	10.764
square meters	square yards	1.196
hectares	acres	2.471
square kilometers	square miles	0.386
<b>Volume</b>		
liters	gallons	0.264
cubic meters	cubic feet	35.314
cubic meters	cubic yards	1.308
<b>Mass</b>		
grams	ounces	0.035
kilograms	pounds	2.205
megagrams	short tons	1.102
<b>Illumination</b>		
lux	footcandles	0.093
candelas per square meter	footlamberts	0.292
<b>Force and Pressure or Stress</b>		
newtons	poundforce	0.225
kilopascals	poundforce per square inch	0.145
<b>Temperature</b>		

To convert Celsius temperature (°C) to Fahrenheit temperature (°F), use the following formula:

$$^{\circ}\text{F} = (^{\circ}\text{C} \times 1.8) + 32$$

## Abbreviations Used Without Definitions

AASHO	American Association of State Highway Officials
AASHTO	American Association of State Highway and Transportation Officials
ACRP	Airport Cooperative Research Program
APTA	American Public Transportation Association
ASCE	American Society of Civil Engineers
ASTM	American Society for Testing and Materials (known by abbreviation only)
FAA	Federal Aviation Administration
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
ITE	Institute of Transportation Engineers
NASA	National Aeronautics and Space Administration
NCHRP	National Cooperative Highway Research Program
NHTSA	National Highway Traffic Safety Administration
RITA	Research and Innovative Technology Administration
SAE	Society of Automotive Engineers
SHRP	Strategic Highway Research Program
TCRP	Transit Cooperative Research Program
TRB	Transportation Research Board

PART 1

**2011 Thomas B. Deen  
Distinguished Lecture**



# James W. McClellan

## 2011 Thomas B. Deen Distinguished Lecturer

James W. McClellan is vice president of Woodside Consulting Group, Virginia Beach, Virginia. His unique career in rail transportation has included service in various capacities for Southern Railway, New York Central–Penn Central, FRA, Amtrak, the U.S. Railway Association, the Association of American Railroads, and Norfolk Southern (NS). Milestones in which he participated include the creation and start-up of Amtrak, the creation of Conrail, the NS merger, and the split of Conrail between NS and CSX.

McClellan retired in 2003 as senior vice president for planning at NS, where his areas of responsibility included rail mergers and acquisitions, sale and abandonment of rail lines, strategic and network



planning, and public–private partnerships. He chaired the NS infrastructure team, which dealt with infrastructure and capacity issues, including the operation of passenger trains. He also provided technical liaison with FRA, the Surface Transportation Board, and state departments of transportation.

For TRB, McClellan has served on specially appointed committees for the studies on landside access to U.S. ports, freight capacity for the next century, and funding options for freight transportation projects of national significance. He is a member of the Committee for Review of the FRA Research and Development Program and the National Cooperative Freight Research Program project panel on Operational and Low-Cost Improvements to Freight Transportation System Performance. He also served on the oversight committee for the Innovations Deserving Exploratory Analysis Program for high-speed rail.



# Railroads and the New Normal

## Impact of Lean and Green on the Future of Railroads

James W. McClellan

Railroading faced a bleak future in the 1960s. At the heart of the decline were several core problems. Freight and passenger traffic was shifting to highways, with the rapidly developing Interstate highway system, and to airlines, with the growing use of commercial jet aircraft. There was an overabundance of railroads and of rail capacity for the diminishing rail traffic. Unproductive labor agreements limited productivity, and economic regulations restricted flexibility. A series of changes over the course of more than two decades produced a more efficient, environmentally friendly—or lean and green—railroading industry. The size and the number of railroads diminished, regulations were eliminated, and more flexible staffing and work rules were instituted with nonunion labor. Better technology also improved productivity. Railroads are now the most financially successful of all transportation modes although immense challenges still lie ahead.

The theme for this lecture was taken from an editorial in *Fortune Magazine* written in the depths of the recession. The thesis was that in the coming years the firms that would prosper had to be efficient—lean, if you will—and environmentally friendly, or green. That got me to thinking about the implications of lean and green for the railroad industry. The easy answer was that railroads were both lean and green already so they were ahead of the game. It is clear—to me at least—that the industry will face some immense challenges as well as some great opportunities.

Before delving into the future, which is always subjective, I would like to spend some time outlining just how far the railroads have come since the 1960s, when bankruptcies and declining market share led many to believe that railroads were doomed to go the way of the horse and buggy.

### FROM RAGS TO RICHES OR THE COMEBACK KID

When I started in the business in the early 1960s, the Interstate Highway System was expanding rapidly, producing huge improvements in automobile and truck transportation. Jet aircraft were coming into service and would proceed to turn airlines from a mode for the elite to transportation for the masses. Railroads were being hammered by the loss of freight and passenger traffic and the future seemed bleak, especially for railroads in the Northeast and the Midwest.

Things did not work out exactly as expected. Airlines stumbled and then stumbled again, and they continued to seek the holy grail of sustained profitability. Autos are the core of our passenger transportation network but they too are less shiny than they were in the 1960s. Their success has brought sprawl, pollution, and often massive congestion. Trucking is the dominant freight mode but, as with the airlines, profitability has been elusive and congestion and pollution are a by-product of their success. More controversial is the impact of large trucks on pavement conditions, throughput capacity, and safety.

Railroads are now solidly profitable and are the most financially successful of all transportation modes. It is certainly the transportation comeback story of the past two decades. After all, Warren Buffett owns a major railroad, the Burlington Northern Santa Fe (BNSF), but has no major investments in either the airline or the trucking industry, though he did flirt with U.S. Airways some years ago.

So how did this turnaround occur and is there anything we might learn from it?

The simplistic notion of the rail revival was that Congress passed a law, in this case the Staggers Rail Act of 1980, and all was right in the world again. Railroads, that story line goes, were overregulated and when the dead hand of regulation was lifted by the Staggers Act, life was good once more. The story fits a fairy tale notion that complex problems can be fixed if Congress just passes the right law—instant gratification through legislation, if you will.

The simple story centered on the Staggers Act misses most of the story. The Staggers Act dealt with some, but by no means all, of the railroads' difficulties. It took well over two decades to implement all the pieces needed for the rail recovery. And while a lot of smart decisions were made, a lot of luck was involved as well.

The decline of the railroads started early in the 20th century. Railroads are highly efficient but they are also inconvenient. Automobiles and trucks are far more flexible and user-friendly, so as the road system expanded rail traffic suffered. World War II provided a respite from traffic losses as both auto and truck transportation was discouraged by fuel and rubber rationing. As soon as the war ended, however, the decline of the railroads continued apace and at an accelerating rate.

By the early 1960s, a full-blown railroad crisis existed, at least in many parts of the nation; some railroads, mostly in the West and South, were never in dire financial straits. Still, the railroad problem, which was focused first in the Northeast and Midwest, ultimately spread from coast to coast as railroads such as the Rock Island and the Milwaukee Road failed and the Chicago and North Western Railway and the Southern Pacific Railroad nearly did. As losses soared, so did deferred maintenance, and with that service deteriorated. And service deterioration caused more traffic losses, setting up a potential death spiral.

---

Woodside Consulting Group, 1332 Holly Point Road, Virginia Beach, VA 23454.  
jimmcclellan@att.net.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 5–12.  
DOI: 10.3141/2261-01

At the time, railroads were prone to blame their ills on government subsidies to other transportation modes as well as rigid, heavy-handed regulation. That was true but was not the core issue. The hard fact was that autos and trucks were a better mousetrap for much of the transportation market. Again, convenience trumped efficiency for the users. Government invested in other modes not as a plot against railroads but because people demanded better roads, or in the case of inland waterways because Congress knew that low-cost barge service would help American farmers and grain dealers. That investment, most notably the Interstate highway system, certainly sped up the shift in traffic to auto and truck but was ancillary to the railroads' decline.

## CORE RAILROAD PROBLEMS

The following core issues affected railroad viability:

- **Loss of traffic.** There was a substantial shift of freight and passenger traffic to the highways and of passenger traffic to the airlines. Some traffic simply stopped moving altogether as industries, especially in the Northeast, closed. Railroads in the West and South fared better; while their share was also declining, the underlying economies were expanding.
- **Overcapacity.** Too much railroad was chasing too little traffic. Management had been slow to adjust capacity to lower levels of demand. While regulation made the exiting of markets difficult, management was often reluctant to admit that traditional markets were gone and were not coming back. Recently, similar evidence of management denial has been observed, most notably in the domestic automobile business.
- **Unproductive labor agreements.** As rail was a heavily unionized industry, wages and benefits were high and work rules limited productivity. Much of the trucking industry (especially the full truckload business most competitive with railroads) was free of such restrictions. The loss of market share from domestic auto producers to foreign-owned transplants with lower labor costs is a similar phenomenon.
- **Huge and rapidly escalating passenger train deficits.** Passenger trains had been a money-losing business for decades, but with the rapid post-World War II decline in passenger volumes, followed by the loss of most mail contracts in the late 1960s, losses ballooned. By the end of the 1960s, passenger deficits threatened even the financially strongest carriers. Regulators eventually allowed many abandonments; the process was both slow and uncertain.
- **Too many railroads.** In a world where trucks could serve any market directly, railroads were balkanized into regional entities. Through service took a lot of coordination of both services and pricing. Different railroads had different goals and objectives; working together was difficult at best and often impossible, at least in a timely fashion.
- **Economic regulation.** Railroads could not raise or lower rates without regulatory approval. Entry and exit also required regulatory approval, as did rail mergers and transfer of lines. Changes in pricing or in the network could take years; one merger—the Rock Island case—dragged on for more than a decade until the whole plan was moot. Almost anyone—customers, employees, unions, competitors, communities—could file objections to a proposed action. The railroad where I started my career spent 3 years and millions of dollars implementing lower grain rates offered in large new covered hopper cars replacing antiquated, inefficient box cars; the Interstate Commerce

Commission (hearing objections from some customers as well as other railroads and barge lines) decided against the Southern twice, until it was finally overruled by the Supreme Court.

## SOLVING THE RAILROAD PROBLEM

Each of these core issues was dealt with and all had to be addressed to put railroading on a viable footing. The process took more than two decades. Here is a quick review of some critical actions taken by both the private and the public sector.

### Passenger Deficit

The passenger deficit was tackled first. The free market choice of simply letting the trains die was rejected and the only way to save even a core system was for intercity trains to be subsidized by the government. It was not a great answer; the system shrank by 50% and has cost billions in subsidies. The myth was that it would operate at a profit at some point; even those of us on the inside knew that was impossible. We reasoned that it was better to have a skeletal system than none at all, and it was certainly good public policy to free freight railroads from most, though not all, of the losses. Decades after it started, Amtrak still relies on below-cost, statutorily guaranteed access to freight tracks. Again, the amount of the subsidy is controversial, but consider just this one caveat: if Amtrak decides to run more service on the BNSF transcontinental mainline to California, it pays nothing for the opportunity cost to BNSF of using that capacity for its own intermodal traffic.

This means that future public policy green initiative actions intended, for example, to divert significant numbers of trucks from the highways runs smack up against competing green initiatives to add passenger service.

Intercity passenger trains were only part of the passenger problem. Commuter systems in Boston, Massachusetts; New York; Philadelphia, Pennsylvania; Baltimore, Maryland; Washington, D.C.; Chicago, Illinois; and San Francisco, California, lost millions of dollars annually. Shifting those losses to the public sector was a slow and costly process that took more than a decade.

### Overcapacity

Overcapacity was tackled first by the U.S. Department of Transportation (DOT) in its efforts to deal with the bankrupt railroads in the Northeast. First, DOT issued a report and focused on the fact too many railroads and too much railroad capacity were chasing after too little traffic. DOT suggested that 25% of the entire northeastern rail network, operated by both solvent and bankrupt carriers, was potentially surplus—a politically correct way of saying “tear it up.” The howls from all constituents were huge and Congress created a new entity to address the issue, the U.S. Railway Association (USRA). In its final system plan, USRA excluded thousands of route miles from the new Consolidated Rail Corporation (Conrail) system, leaving that mileage to be abandoned or operated with government subsidies. Congress had no ability to tweak the plan, only the ability to reject it and seek a new approach. Critics of the plan lacked the votes to force such an outcome.

Over the next 15 years, line rationalization became a major corporate initiative on all major railroads. In the Midwest, which followed

the Northeast with a series of bankruptcies and sick railroad problems, DOT and FRA played a critical role in stimulating rationalization, in part by tying many low-cost loans and grants to line rationalization efforts but perhaps more importantly by supporting the Milwaukee Road trustees in their efforts to shrink to a core operation, thus avoiding a Conrail-West solution. DOT urged the Rock Island trustees to do the same, with less success. Even the solvent railroads, taking a cue from the success of Conrail and others, undertook programs of their own. In the 1980s, I led the effort at Norfolk Southern that shrank the system by 6,000 mi, a third of its system, through transfers to short-line carriers and by outright abandonment.

### Too Many Railroads

The problem of too many railroads was already being addressed when the railroad crisis hit. Mergers were a business strategy from the early days of railroading but the modern merger movement began in the 1950s with the merger of the Norfolk and Western Railway and the Virginian Railway. The merger of the Pennsylvania Railroad and the New York Central Railroad and its later failure put a pall over the effort. Nonetheless, with the creation of the Burlington Northern in 1970 and Conrail in 1976, mergers were considered a necessary part of the perpetual (or so it seemed by then) railroad problem. The Staggers Act played a major role in the post-Conrail mergers. The Act made viability of the industry an important goal and tilted the playing field toward protecting competition, not competitors. The result was an Interstate Commerce Commission that was far more willing to allow mergers. CSX Corporation started the ball rolling despite efforts by Norfolk and Western and Southern to extract punitive conditions. The dominos continued to fall until the industry settled on a six-carrier structure: two major carriers in the East, two in the West, and two based in Canada (each of which has substantial U.S. operations).

The final big event in the merger puzzle, the split of Conrail between CSX and Norfolk Southern, came more than 20 years after Conrail was created as a monopoly and 40 years after the Norfolk and Western–Virginian merger. The Conrail split is especially noteworthy; the merger of the Pennsylvania and the New York Central Railroads created a competitive imbalance in the Northeast. USRA tried but failed to resolve the issue and Big Conrail was created, making a bad imbalance even worse. The split, a private sector–funded solution, restored the competitive balance that had been lost more than 30 years earlier.

### Economic Regulation

Economic regulation was substantially changed by the Staggers Act. Railroad technology and economic advantage favor large, point-to-point movements: the less intermediate handling, the lower the costs; the heavier the load per car and per train, the lower the costs; and the longer the haul, the lower the costs per mile. Railroad economics of scale and density—elusive before the Staggers Act—finally became meaningful.

The flexibility provided by the Staggers Act—in particular its legalization of pricing and service contracts between railroads and their customers—made it possible to quote lower prices for volume movements or for long-term commitments of traffic. It became easier to close unprofitable routes and junctions, thus improving lane density and reducing switching costs. Some prices were raised but

mainly railroads reduced prices to give incentives to customers to load heavier and ship on direct routes. Increasingly, prices began to reflect the economics of the business.

Deregulation did not come easily or even quickly. Some railroads opposed the Staggers Act and many were slow to implement changes.

The Staggers Act did not change everything. For example, abandonments were still regulated under the old rules, but transfers of lines to new railroads were liberalized. Still, the declared purpose of the Staggers Act was to maintain and enhance the viability of carriers and not just to provide a level playing field for all competitors and all customers. Arguments involving the right balance between viability and efficiency and competitive access for customers continue to rage to this day.

As an important footnote, Conrail gave the greatest impetus to passage of the Staggers Act. Launched in 1976, Conrail continued to lose money and required annual operating subsidies. Congress grew tired of the payments and Conrail argued that the deficit could be eliminated if it could restructure its operations and pricing.

### Unproductive Labor Agreements

Dealing with unproductive labor agreements was another area where change came slowly. USRA essentially ducked the problem when Conrail was formed; it was too hot a political issue. Later, however, government played an important role in introducing change. When the unions went on strike against the Rock Island Railroad in 1978, the Carter Administration allowed the railroad to be liquidated—much to the surprise of the unions, which had expected a Conrail-type bailout. Next, the new Reagan Administration, having already fired the air traffic controllers, threatened to liquidate Conrail unless its costs were brought into line. Conrail used that threat to reduce its train crews from four or five to three.

The Interstate Commerce Commission finally took a liberal position on transfer of rail lines to short-line and regional carriers under Staggers Act principles. One divestiture after another was approved and all were made to low cost, generally nonunion companies with flexible work rules. The threat of more transfers put pressure on the unions to adopt more flexible work rules and the major carriers continued to push for change.

A major fork in the road came with Presidential Emergency Board 219. In an arbitration dispute over crew size, the board sided with the Chicago and Northwestern and found that a two-person crew on through freight trains was reasonable. The union planned to strike and that would have shut down much of the Chicago commuter network. Congress, about to leave for summer recess, simply forced adoption of the Presidential Emergency Board decision on the Chicago and Northwestern unions. After that, the other major carriers were able to negotiate more flexible staffing and work rules.

### Technology

Technology played a huge role in improving productivity. It would take an entire evening to list all the changes—many of them incremental—that made a difference. Higher-quality rail meant a longer time between replacements. Mechanization of track maintenance produced greater productivity as well as less downtime for maintenance. Less downtime meant more time to run trains, further improving the productivity of line-haul crews.

Centralized traffic control makes it possible for trains to take sidings without stopping the train and throwing a switch; thus, trains move faster and crews are more productive. The development of remote control locomotives permitted the use of one-person crews in yards. Trackside monitoring devices have eliminated many en route inspections and have greatly reduced the number and severity of derailments caused by bearing failures or broken wheels or loads that were either too high or too low for clearances of the route. Improved dispatching models are coming into use that will increase the capacity of existing lines.

The productivity numbers are impressive. In 1965, ton miles per employee numbered 1.1 million. By 2006 and before the great recession, ton miles per employee numbered 10.6 million. A somewhat fairer measure is to compare the results from 1990 and 2006; by 1990, passenger service employees had largely been transferred to passenger entities and most of the transfers to short lines had been made. The results were still impressive, ton miles per employee increased from 4.8 million to 10.6 million.

Unlike many manufacturing jobs, the work was not outsourced to another country. Technology and changes in work practices drove the railroad result. As we think about where the jobs have gone, the results for the railroads show just how powerful technology can be. All the technology came with a massive price tag, which is why many in the railroad industry view efforts by Congress to roll back the clock on railroad regulation with such alarm.

### Loss of Rail Traffic

The loss of rail traffic that so devastated the northeastern railroads was halted and even reversed. Two major sources of new traffic ultimately produced an explosion of rail traffic, though the western carriers benefited far more than the eastern carriers. One source, the movement of low-sulfur coal from Wyoming and Montana was the outcome of government action (and once again, government action proved important); the Clean Air Act of 1970 and Clean Air Act Amendments in 1990 mandated a reduction in sulfur emissions from coal-fired power plants. A utility had a choice of installing scrubbers or burning low-sulfur coal. Most opted for the latter and today western coal moves all the way from Wyoming to New Jersey and Georgia.

The 500-mi hauls from Appalachia were replaced with 1,500- or 2,000-mi hauls. Not only that, but the British thermal unit content of western coal was much lower than that of eastern coal, so it took a lot more tons to provide an equal amount of heat. The bottom line was a lot of tonnage moving long distances. It was a perfect commodity for rail technology: heavy loading and usually long hauls in unit trains.

Ton miles exploded, driving rail market share up. Still, traffic was not diverted from the highway, which shows how misleading broad market share data can be as a measure of truck versus rail competition. Often a change in share data reflects not a shift between modes but rather whole new sources of traffic that simply did not move before.

The other major engine of growth was intermodal, especially international intermodal. Some railroads experimented with containers in the 1960s for domestic intermodal. The game-changing technology, however, was the container ship and the double-stack container car. By going up, a train could carry almost twice the payload in a given length. The double-stack car, which is pretty much the standard for intermodal services today, was designed and

built by a consortium of a steamship company, a railroad, and a car builder.

Big ships serving major ports fed trainloads of cargo, which then moved long distances by rail. Hauls of 2,000 mi or more were, and are, common. It was a fast way to rack up ton miles. And intermodal provided another efficiency benefit. Often there was no gathering function at all; traffic was loaded at the port directly to rail cars. And the gathering and distribution that was done used contract dray operators: nonunion labor with flexible work rules and the ability to come into and go out of the labor pool as conditions warranted.

Much of the improved efficiency that railroads achieved came from a change in traffic mix; traffic moving in trainload quantities assumed a greater share of the rail traffic base while the higher-cost, single-car business declined in relative, though not absolute, importance.

Critically, all this growth required massive investments in infrastructure. Tunnels were enlarged and bridge profiles were modified to accommodate double-stack trains. Intermodal terminals were built, often with price tags well in excess of \$100 million. New rail lines were built and thousands of miles of existing trackage was upgraded with better signals, additional tracks, and heavier rail and ballast. Now, the Powder River Basin hosts about 120 trains a day, and three or four can often be seen moving at the same time.

Again, give credit to the Staggers Act for much of this increased traffic. Without the ability to make long-term contract rates, railroads could not have raised the capital needed for these massive investments.

Market share, which was at 30.3% in 1989, rose to 43.3% in 2006 and the share of ton miles between rail and trucks has stabilized; the number for 2006 was 43% for rail and 30% for trucks. Ton mile market share, often used by the rail industry for comparative purposes, understates the importance of trucking, for trucks handled 75% of all freight tonnage. Trucking just does not move as far as rail traffic.

All in all, it has been a great success story. And while the private sector played a critical role, the public sector deserves a lot of credit. Government provided the funds needed to lift passenger train losses off the backs of the freight railroads. Government provided the grants and loans to bankrupt and marginal carriers that helped them rebuild their infrastructure.

The turnaround could not have been accomplished without passage of the Staggers Act and, more importantly, a more enlightened regulatory process that recognized that efficiency and viability were an important part of the equation.

In providing financial resources, government practiced tough love. Fundamental change was the price of government funding. The railroad story gives us some hope that perhaps the current intervention in the auto industry will have a happy ending.

### LESSONS TO BE LEARNED

There are some lessons that can be learned:

- It takes a long time to get deeply in trouble and it takes a long time to fix the problem. Just passing a law does not provide instant gratification. So fixing General Motors or the housing industry or health care is going to be a hard slog and progress will often be erratic—two steps forward and one step back. Patience is not exactly an American virtue, and that will be a problem as we tackle problems far more complicated than failing railroads.

- Ignoring economic realities and market forces is a huge mistake for societies as well as families. Amtrak was created on a political compromise that created the myth that passenger trains could make money. Conrail was created in a more objective process and the upfront capital to rebuild the system was recognized and allocated.

- Restructuring is a painful process. Essentially, the railroads had to reduce costs to compete. In the process, hundreds of thousands of employees lost well-paying jobs and thousands of communities lost rail service.

- Government can be a powerful catalyst for positive action but that may be achieved only when there is a crisis.

- In the end, economics trumped politics. There were many arguments in favor of keeping the status quo. Government, however, was unwilling to pay for the long-term costs of maintaining the status quo. (Regulation, is the reverse, politics trumping economics.)

- Congress ought to be written out of any detailed solution. The tough love of the final system plan was sustained because Congress had no ability to tinker with the plan.

## GOOD NEWS FROM THE GREAT RECESSION

Just how far the railroads have come was underscored by their performance in the great recession. Traffic levels began to get a bit soft for both railroads and truckers in 2007 and most of 2008. Then, in the fourth quarter of 2008, the bottom fell out of the economy. Rail traffic collapsed, down 20% or more on most railroads. Based on historical precedent, a supposedly high-fixed-cost industry should have been crippled by such a huge and rapid loss of traffic.

That scenario did not happen. Railroads adjusted and they adjusted quickly. Earnings declined but did not disappear as operations were rapidly brought into line with diminished traffic levels. There was a crisis in the auto sector, a crisis in the construction sector, and a crisis in the domestic auto sector.

There was, however, no railroad crisis.

Railroad performance in the great recession was in marked contrast to the 1960s when financial collapse was imminent for many companies and the merest dip in revenue led to disaster. In my view, these are the factors that made a difference:

- Now there are six major, cohesive networks. Train operations can be changed over large networks on almost a real-time basis. In the 1960s, negotiating service changes with a dozen or so connecting carriers made any rapid response to changes in volume almost impossible, at least on a timely basis.

- Now powerful network models allow railroads to reprogram their networks rapidly and with confidence in the outcome. In earlier days, changes were made by the seat of the pants; changes were made and then observed to see how they turned out. Often, they turned out badly and it was back to the drawing board.

- Now railroads have moved a lot of assets and people to other parties. A substantial portion of the car fleet is owned by customers. Much of the gathering function is provided by short-line carriers or, more importantly given the importance of intermodal services, by third-party draymen. As volumes plummeted, so did equipment rents and payments to draymen and short lines.

- Current managements are simply more aggressive now than those in the past. Railroading used to have a semi-public-utility attitude, and decision making was slow and bureaucratic. That is no longer the case.

This year, as traffic is increasing, railroads have added a few employees but generally handle more business with fewer resources than prerecession. Trains are longer, older power remains stored, and marginal facilities closed. Recent numbers show that rail traffic is up by 28% but the number of train and engine crew starts is up only 6%. This performance, which has been repeated throughout much of the private sector, is a reason why unemployment remains so high. As an aside, just think about all those marginal auto plants that were closed as production was concentrated at the newest and more automated facilities. Lots of jobs disappeared and will not come back.

In short, railroads have been reinvented and their performance in this great recession proves that the transformation can produce solid results in good times and bad.

## WHERE DO WE GO FROM HERE: THE IMPACT OF LEAN

Now I will turn to the future. Looking into the future is certainly a daunting task, and there are about as many forecasts of the future as there are economists, pundits, and politicians making them. So what follows is speculation, though informed speculation, I think. I worked on restructuring railroads for much of my career, expanding into markets where it made sense and exiting markets that looked weak. By necessity, I spent a lot of time thinking about the future: how much would be moving, where it would be moving, and so on. One tough thing about railroading is that you cannot pick up the tracks and move them if a market shifts.

As I write this, the economic landscape is fairly bleak. Outright disaster has been avoided but most economists now predict a long, slow recovery. The net worth of households is about where it was a decade ago. Consumers drove the old economy but it turned out they did it with easy credit; it was a false prosperity funded with unsustainable debt. It is difficult to see where consumers, facing stagnant wages at best and unemployment at worst, are going to find the bucks to create a surge in demand. In short, the country has become far leaner, though the choice was not voluntary.

An enforced lean economy has some real consequences for transportation. Transportation exists to move things, stuff. When people buy less there is less to move. In fact, the great recession has already wiped out about 10 years of growth in the rail carload business. Nor have truckers escaped; they need a number of good years to return to the traffic levels of 2006.

For the transportation professional, the great recession and its aftermath will be felt for years, if not decades, to come. A slow, uneven recovery seems the most likely scenario: two steps forward and one step back. Now the mortgage documentation debacle and all it portends on the recovery dominate the news: one step backward.

Things might be a lot better than they look right now. It is always a bit dangerous to forecast from either the top of an economic cycle or the bottom. Assuming we will have a slow recovery, certainly a plausible thesis, there are implications for transportation. We, as transportation professionals, may be forced to rethink some of our basic beliefs, including the following.

### Growth Rates

A leaner future means that transportation volumes may not grow at the rates that we saw in the recent past. There will be growth as the

population expands, and people do have to eat and will continue to consume. However, the hyperconsumption that characterized the period before the great recession is likely a thing of the past. Think about a leaner future. The population is going to increase, but the amount of freight consumed per capita is likely to decline.

We should rethink capacity needs whether for highways, airports, railways, or ports. "Build it and they will come" just may not be true anymore; it does not seem prudent to most business people and probably no longer works for a public sector that is stressed for funding. Just look at the unprecedented decline in housing and the huge inventory of unsold housing that will slow suburban sprawl in coming years, and thus the need for more roads.

### Availability of Government Funding

A leaner future will put a lot of pressure on transportation funding. Transportation professionals tend to get caught in their own silos. Thus, we argue that the transportation infrastructure needs are ever mounting and must be funded. Those in education, health care, and the military are also caught in their own silos and are certain that their needs must be funded. And we all know that government, at all levels, is fundamentally out of money. Further, government has limited means to raise revenues in a depressed economy; people might have accepted increased sales taxes or gasoline taxes when times were good but are unlikely to do so in today's environment.

Caught somewhere between meds for Grandma and retirement for Grandpa and schools for Sammy and no new taxes and covering all the public sector unfunded pension plans means that public sector finances are going to be a mess even when the economy rebounds. The recent cancellation of the trans-Hudson rail tunnel by the governor of New Jersey is not the last of the bad news for transportation.

### Tilt Toward Efficiency

If this scenario plays out, the transportation industry will have to find ways to do more with less, which means making better use of the infrastructure that we have. Transportation decisions have been driven by convenience, not efficiency, and that is the inconvenient truth. Convenience will still be important, but the future world is likely to tilt more toward efficiency. We are seeing this borne out by the rapid growth of domestic rail intermodal, where major truckload carriers have embraced intermodal as a way to move traffic at less cost than going over roads. In the process, tens of thousands of truck drivers have been displaced, which does nothing to solve the unemployment problem.

### Tilt Toward More Private-Sector Financing

If the public sector no longer has the wherewithal to fund necessary transportation improvements, it is possible that the private sector (which has a far more robust balance sheet at this point in time) will step in. Public-private partnerships are going to be a lot more common but the old model of augmenting private capital with a big dose of public funds will shift. The public dose, by necessity, will be far reduced.

Reduced consumption and a stressed public sector are not the only aspects of a lean future. Products are being downsized. Consider

houses and motor vehicles. Houses, to the extent any are even being built, have been downsized simply to bring their price in line with what people can actually afford. There will be more multifamily dwellings as renting replaces owning for many people.

Smaller dwellings mean less stuff to move, from construction material to carpeting, and less coal to burn to provide the power for heating and cooling. While these changes will be at the margin, they will limit future growth.

With regard to motor vehicles, they are an important source of freight traffic. Steel and plastic and various parts are moved and the finished vehicle is moved to its destination. Lean means that fewer vehicles will be produced; 2010 looks like it will come in at less than 12 million units, up substantially from the depths of the great recession but well below the peak production years of the recent past.

Not only are fewer units being produced, they are on average smaller and lighter than was the case just a few years ago. Smaller means less steel, less plastic, and less carpeting and it means that more vehicles can be loaded on a single railcar. It also means less fuel use and lower emissions. From a ton-mile-generation standpoint, however, railroads would be better off if everyone bought a really big sports utility vehicle.

Lean does not end there. Walmart, for example, is making a concerted effort to reduce its carbon footprint by forcing its suppliers to redesign packaging so that more goods can fit into a trailer or container.

In short, lean means that economic constraints will reduce the amount of stuff that is bought and that stuff will be smaller and lighter. It is good for society in the long term but still is something that will limit the growth of freight transportation.

### IMPACT OF GREEN

Another mega trend that will affect transportation is concern for the environment, the green part of a lean and green future. The arguments about climate change will ebb and flow and it seems unlikely that we will get any comprehensive energy legislation. Still, concerns will be felt in local and state ordinances and regulations, while any progress at the national level is increasingly problematic. Those concerns, however, will not stop a multitude of state and local initiatives because concern about the environment is not just part of a liberal agenda.

Railroads advertise themselves to be the green mode, and the arguments are fairly compelling. Deeper analysis shows that environmental concerns will create some substantial downside risks for railroads. A few of those risks are discussed below:

#### Coal

Coal is first in tonnage and usually first in profitability on most railroads and comes in second in total rail revenues, just behind intermodal. Environmentalists have painted a big bull's eye on coal-fired power plants.

Coal is not going away anytime in the next several decades. There is too much investment in coal-fired generation plants, and wind and solar are more expensive. The poor state of the economy will keep coal in business for longer than it would were the economy stronger. Obtaining higher utility rates in these economic times for costly solutions such as nuclear or solar or wind or coal sequestration is a

difficult proposition. Getting public sector subsidies will be tougher as well.

That said, coal is going to be under a lot of pressure. It is difficult to get a new coal-fired generating plant approved. As older coal plants are retired, production will shift not to exotic sources such as wind but to gas-fired plants. First, gas is increasingly plentiful. Second, the carbon dioxide emissions have half the impact of coal. Third, gas is increasingly competitive in cost, which means the environment is improved without everyone's utility bill soaring out of sight and it moves by pipeline, not by rail.

As discussed earlier, western coal was a major engine of growth for railroads. Coal will continue to be important but will not provide the growth that it did in the past.

## Downsizing

There will be less stuff to move. We have already talked about the downsizing of homes and motor vehicles and the use of more efficient packaging as part of a trend toward a leaner future. Green reinforces that downsizing and even if the economy makes a huge recovery, it is hard to see gas guzzlers coming back in large numbers simply because government mileage standards will keep manufacturers focused on somewhat smaller, fuel-efficient vehicles.

Green, coupled with community concerns, will continue to cause delays and increase costs for new transportation projects. These delays will be a special concern as we try to stretch what may well be scarce transportation dollars. The double whammy of limited funding and environmental and community concerns are likely to prove highly toxic for large-scale highway projects. Railroads, however, will have significant problems as well, as I will discuss in a moment.

Given the twin forces of lean and green, how will the freight railroads evolve? I see the next decade unfolding as follows.

## Carload Business

The carload business, including coal, will grow with the economy and will stay below historic levels until the economy rebounds. As I said earlier, I think that is going to be years not months.

## Intermodal

Intermodal, more specifically domestic intermodal, will be the engine of future growth. In recent months, intermodal has rebounded to levels that are almost back to the peak year of 2006. International intermodal will come back as the economy recovers but domestic intermodal will be the real story.

Domestic intermodal will be pushed to some extent by continued constraints on highway capacity. Public sector financial woes as well as green resistance to many highway projects will make new highways a tough sell. This is not altogether good news for rail intermodal as it is very dependent on the urban-suburban highway network, and the same green concerns will make building on new intermodal terminals time-consuming, expensive, and in some cases impossible. As a further caution for those who long to take trucks off the road, most truck movements are less than 300 mi, distances where intermodal technology has been unable to compete.

## Capacity

Capacity will not be a serious issue but a number of choke points will remain, especially as the intermodal business ramps up. The decline in traffic during the great recession has given the railroads some breathing room, and slow growth will continue to provide some relief from capacity issues. At the same time, sophisticated dispatching models and improved real-time information on train location, which can improve line capacity, will mitigate some of the capacity issues.

Adding capacity for intermodal growth will be a challenge, especially in major terminal areas where slower speeds, passenger and commuter train operations, and crossing with other railroads cause congestion. Major rail hubs, like major highway hubs, are where congestion is worst. It is at those same locations where adding capacity is both costly and time-consuming and often runs up against community and environmental concerns. Not only is track capacity hard to add in major urban areas, so too is finding places to build intermodal facilities. An intermodal terminal facility operating 24/7 is not anyone's idea of a good neighbor. It is clear that intermodal will be the engine of growth in the future, but it is not clear that adding the needed capacity can be accomplished in an efficient and cost-effective manner.

Finding the funds to build needed capacity will be a challenge, especially in light of the unfunded mandate the railroads face to install positive train control. That unfunded mandate will drain about \$15 billion in capital from the industry, funds that could have bought a lot of additional intermodal terminal capacity, among other things.

## Railroad Funding

Freight railroads will continue to tap public funding for some projects, but the availability of such funding is uncertain at best. One thing I learned when dealing with the northeastern railroad crisis of the 1970s is that freight transportation in general and railroads specifically are not high on a politician's wish list.

If the public money dries up, it is not the end of the world. Compared with overall capital investment, it is not a significant number. Railroads will be able to raise funds in the private sector for any project that makes economic sense. Looking at the demands on public sector budgets, I would certainly rather be trying to add railroad capacity than highway capacity. Simply stated, railroads are solvent and the government is not.

I have not dealt with the future of rail passenger services. That subject deserves a complete lecture in its own right. I do, however, have a warm spot in my heart for passenger trains; I was present at the creation of Amtrak and was one of its first employees. I think there is a role for passenger trains in selected markets and certainly they have proven their worth in the Northeast Corridor and in California. This country has never shown any inclination to provide long-term funding for intercity passenger trains. The recent passenger rail initiatives are more of the same in that regard and, given the state of public sector balance sheets, especially in the states where passenger rail makes the most sense, where are states going to find the reliable, long-term funding needed to provide the ongoing operating subsidies required? I see no clear way forward for rail passenger service, no matter how passionate its supporters might be. Again, the pressures of meds for Grandma will trump public sector budgets.

## SUMMING UP

The nation now faces an extraordinary set of challenges. It sometimes seems that there is no end to bad news. That, on a much smaller scale, was what railroading looked like in 1960. Traffic was often declining, trains were coming off, maintenance was being deferred, and deficits were mounting. The then current government regulatory mechanism was dysfunctional and at odds with the existing market and economic reality of railroading.

Still, the problems were addressed, though long after they had become a crisis. Solutions were found and for the most part those solutions worked. The key to that success was to deal with the world as it was and not the world as we wished it to be. And transportation professionals in both the private and the public sector guided the effort.

Going forward, the railroads are well positioned for success. The physical plant is in the best condition of my long career. The people

running the business are smart and more importantly have learned how to deal with a rapidly changing marketplace. The key will be fact-based decision making and not responses based on would have been, could have been, should have been.

I am confident that the railroads will get it right.

Finally, there is a message in all of this for the broader transportation community. Resources are going to be scarce. Transportation, and especially freight transportation, is simply going to have to compete with all sorts of other priorities. It will be up to us, as the professionals, to define what needs to be done and what, realistically, is likely to be funded. The entire transportation community is faced with the same sort of hard economic choices that the railroads faced for more than three decades.

We live in interesting, and demanding, times.

---

*The Rail Group peer-reviewed this paper.*

**PART 2**

**Railways 2011**



# Assessment of Wider Economic Impacts of High-Speed Rail for Great Britain

Daniel J. Graham and Patricia C. Melo

This paper considers the scope for wider economic impacts of agglomeration-based effects in the context of high-speed rail. The paper provides new evidence concerning the distance decay of long-distance travel and on the basis of such estimates assesses the potential order of magnitude of agglomeration benefits from improvements to high-speed rail transport in Great Britain. A new methodology assumed constancy of trip decay with respect to equal average travel times but variability with respect to distance. The expected outcome was that improvement in travel times would reduce the extent to which travelers perceived distance as an obstacle to interaction. Although urban economic theory did not preclude the existence of agglomeration benefits across interregional distances, the empirical evidence suggested that the order of magnitude of agglomeration benefits corresponding to 25% and 50% increases in travel speeds was small.

This paper considers the scope for agglomeration-based wider economic impacts (WEIs) in the context of high-speed rail in Great Britain. The current Department for Transport (DfT) methodology for the assessment of agglomeration benefits is based on a theoretical model of the urban economy that essentially emphasizes intraurban drivers of productivity. Firms in larger cities are more productive because transactions in input and output markets are rendered more efficient through close proximity to a larger scale of economic activity. Applied to the appraisal of transport investments, the underlying assumption is that improved connectivity within a city will compound the benefits of agglomeration by making the spatial economic transactions even more efficient.

High-speed rail will fundamentally change connectivity between cities rather than within cities. The key task of this paper is to consider whether the agglomeration arguments used to justify the existence of WEIs for intraurban schemes can be extended to inter-urban investments. Essentially, the key questions are: Could high-speed rail investment give rise to agglomeration benefits, and if so, how substantial might these benefits be?

This paper makes two key contributions. First, it generates new empirical evidence on long-distance commuting and business movements between and within British labor markets. The second contribution includes the development and implementation of a methodology for the assessment of the potential order of magni-

tude of agglomeration benefits associated with improvements in long-distance travel.

The empirical work undertaken in this paper is intended to indicate only general orders of magnitude of agglomeration benefits that may arise from improvements in long-distance connectivity. It should certainly not be taken as a definitive or exact statement on the possible wider benefits of high-speed rail.

The structure of the paper is as follows. The first section discusses the theory of agglomeration economies and reviews empirical evidence on the magnitude and spatial scope of agglomeration effects. The second section relates the theory of agglomeration to transport interventions and considers whether the arguments for agglomeration benefits still hold in the context of high-speed rail. The third section presents empirical evidence on the distance decay of intracity and intercity commuting and business flows. The fourth section uses estimates of distance decay gradients to indicate potential orders of magnitude of agglomeration benefits associated with improvements in long-distance travel. The last section draws together the main conclusions.

## SPATIAL STRUCTURE OF AGGLOMERATION ECONOMIES

### Theoretical Foundations for Productivity Benefits from Agglomeration

At their broadest level, agglomeration economies occur when agents (i.e., firms or workers) benefit from being near to other agents. Nearness can involve physical proximity, but transport and communications play a crucial role because, in most contexts, speed and low costs in transportation and communication provide a direct substitute for physical proximity.

The discussion that follows is specifically concerned with production agglomeration economies, which derive from proximity between firms, and the sources of those agglomeration economies: workers, other firms, and other facilities. Because it is important to understand what mechanisms drive production-related agglomeration economies, a brief overview of these issues is presented next.

Production agglomeration economies usually mean that the productivity of individual firms rises with the overall amount of activity in other nearby firms, or with the number of nearby workers or consumers. The literature traditionally emphasizes three sources of agglomeration economies, roughly following three examples given by Marshall: linkages between intermediate and final goods suppliers, labor market interactions, and knowledge spillovers (1). Input-output linkages occur because savings on transport costs mean that firms benefit from locating close to their suppliers and customers.

---

Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College of Science, Technology, and Medicine, London SW7 2AZ, United Kingdom. Corresponding author: D. J. Graham, d.j.graham@imperial.ac.uk.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 15–24.  
DOI: 10.3141/2261-02

Larger, denser labor markets may, for example, allow for a finer division of labor or provide greater incentives for workers to invest in skills. Finally, knowledge or human capital spillovers arise when spatially concentrated firms or workers are more easily able to learn from one another than if they were more spread out.

An alternative taxonomy that sheds more light on the underlying mechanisms is provided by Duranton and Puga, who classify the sources of agglomeration economies as sharing, matching, and learning (2). Sharing refers to the sharing of indivisible facilities, intermediate suppliers, workers, and consumers by firms, which reduces fixed costs, allows specialization, and allows firms to share risks. Matching benefits are usually discussed in terms of the benefits of having a dense pool of workers in close proximity to employers. This means that it is easier for different types of workers and employers to find each other, leading to more productive job–worker matches. Learning refers to the transfer of information, knowledge, and skills. Even in a world of fast communication technologies, close connections between people and firms provide more opportunities for learning and face-to-face contact, which facilitates knowledge exchange and transfer of skills.

The empirical implementation involved in the estimation of productivity benefits from agglomeration economies is described elsewhere (3, 4). Briefly, this involves estimating the statistical relation between the economic output of firms and the degree of agglomeration experienced by the respective firms. This is usually done by fitting a production function in which both private inputs (e.g., labor, capital, materials) and public inputs (e.g., local goods, agglomeration) explain the variation in the level of economic output of firms. Considerable evidence exists on the productivity benefits from agglomeration economies. Researchers tend to agree on positive productivity gains from agglomeration economies, but there are substantial differences in the size of the estimates both across and within studies (5); Melo et al. provide a literature review (6). Figure 1 shows the histogram of 531 statistically significant elasticity estimates of urban agglomeration economies obtained from 33 studies. There is considerable variation in the magnitude of the elasticity values, even among the positive estimates. The estimates vary between  $-0.570$  and  $0.658$ , and the unweighted mean (median) is  $0.057$  ( $0.044$ ): this means

that doubling the size of a given region (typically measured in terms of population, population density, or market access) is associated with an increase in economic output of  $5.7\%$  ( $4.4\%$ ). (The list of studies can be obtained from the authors on request.)

### Transport Investments and Spatial Scope of Agglomeration Economies

Transport improvements can increase the strength of agglomeration economies to the extent that they increase connectivity within the spatial economy. Reflecting on the mechanisms described above, it is clear that agglomeration economies depend crucially on the flows of goods, people, or information between locations. Therefore, the geographical scope of agglomeration economies will depend on the rate at which these flows decrease with distance.

A key issue in understanding the spatial scope of agglomeration economies relates to the construction of the agglomeration term itself. This needs to be a variable that represents the potential opportunities for a firm to benefit from the agglomeration mechanisms in its locality. Second, locality must be defined.

The standard setup is to define agglomeration ( $A_{it}$ ) as an aggregation of workers, firms, or population in the geographical neighborhood of each firm  $i$ . Consider employment as an example. Locality is then defined either by using predefined statistical or administrative zones or, more generally, by aggregating employment with higher weights applied to locations close to firm  $i$ , and lower weights to locations farther afield. This type of agglomeration index has the general structure

$$A_{it} = \sum_{j \neq i} a(c_{ijt}) z_{jt} \quad (1)$$

where weights  $a(c_{ijt})$  are decreasing in the costs or time  $c_{ijt}$  incurred in moving between places  $i$  and  $j$ , and  $z_{jt}$  is the variable being aggregated to create the agglomeration index. Graham (7) refers to  $A_{ijt}$  as effective density, defining  $z_{jt}$  as postcode sector–level employment, setting cost ( $c_{ijt}$ ) as the straight-line distance between postcode sectors ( $d_{ijt}$ )

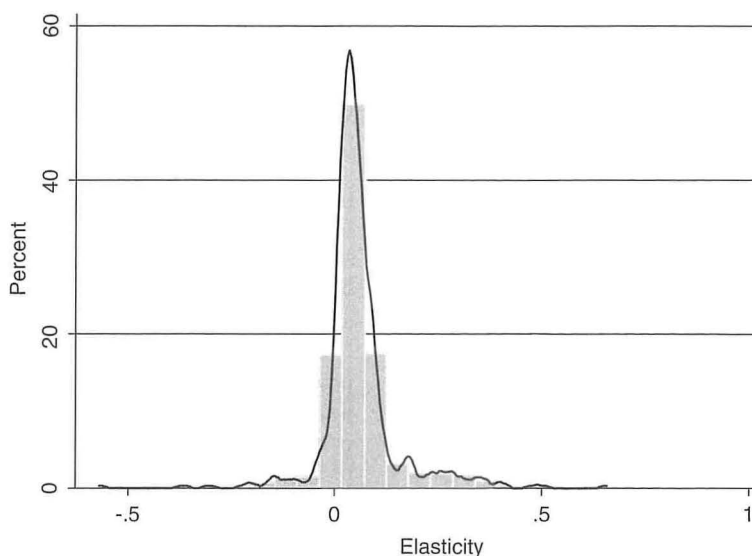


FIGURE 1 Histogram of urban agglomeration elasticities.

and imposing an inverse-distance weighting system (referred to in the literature as gravity based, after Newton):

$$a(c_{ij}) = d_{ij}^{-\alpha}$$

and

$$A_i = \sum_{j \neq i} d_{ij}^{-\alpha} z_{ij} \quad (2)$$

Thus, the effective density measure incorporates the two components that capture the amount of agglomeration experienced by a firm located at site  $i$ : the quantity of employment in another location  $j(z_{ij})$  and the connectedness of site  $i$  with site  $j(d_{ij})$ . The parameter  $\alpha$  is assumed greater than 0, such that employment at place  $j$  has less and less potential influence on a firm at site  $i$  as the distance between  $i$  and  $j$  increases. The larger the value of  $\alpha$ , the more rapidly the potential effect of employment diminishes with distance  $d_{ij}$ . For example, if  $\alpha = 1$ , then the weight attached to employment decays inversely with distance (employment 10 km away from a firm has 1/10th the effect on effective density as employment 1 km away).

The market potential measure described above is identical in form to accessibility indices commonly used in transport analysis (8, 9). One important difference, however, is that market and population potential indices typically use straight-line distances to capture the relative spatial separation between locations, whereas transport accessibility indices generally use distance or travel times along existing transport networks. These distances and times are sometimes converted into generalized transport costs by using estimates of the monetary value of travel time, fuel costs, and so forth, derived from other sources; see, for example, Combes and Lafourcade (10) and Graham (7).

Accessibility, market potential, population potential, or effective density measures based purely on geographical distance depend only on the amount of surrounding employment and how far away that employment is. This may, however, not be the most useful way to evaluate the effects of transport improvements that bring firms or workers (or both) closer. Distance is simply a proxy for the transport costs or travel time separating two locations. Given a fixed transport infrastructure and a fixed transport policy regime, a distance of, say 10 km, has a corresponding (average) travel time or travel cost. To work from a proposed transport improvement to a change in effective density, the expected reduction in travel times or travel costs in each direction must be converted into an equivalent reduction in distance. For example, if a given transport improvement reduces travel costs to the east of a site by 20%, then the new effective density at that site will change in a way that is equivalent to moving employment to the east 20% closer (i.e., effective density with distances in the direction of the transport improvement reduced by 20% must be recalculated).

The more direct way to incorporate transport costs or times into estimates of local economic mass is to base these estimates on existing transport costs or times rather than geographic distances. In this case, local employment counts are aggregated by using a penalty that increases with travel costs or times rather than simple distance. It is then easy to see how to convert a policy-induced change in travel costs or time into a change in accessibility. The drawback of this approach and the reason Graham (7) uses straight-line distances in effective density calculations rather than network distances, times, or costs is that the existing transport network and service are in part dependent on transport demand, which is in turn dependent on the level of economic activity and productivity in a given location. There is thus a risk of inferring that closer connection to employment

increases productivity, when it is productivity that has encouraged closer connections through development of the existing transport network.

## Empirical Evidence Concerning Spatial Scope of Agglomeration

One of the key points emerging from empirical research on agglomeration economies is the insufficient understanding of the pattern of spatial decay of its effects, both the overall effects and the effects associated with the different mechanisms determining agglomeration. The investigation on the spatial decay of the benefits from spatial concentration is a fairly recent research topic and only some papers provide evidence.

Overall, the evidence suggests that agglomeration effects tend to be stronger within short-distance ranges, beyond which their magnitude decreases sharply. Rosenthal and Strange find that worker productivity (measured by wages) in the United States increases by 1.5% to 2.14% for an additional 100,000 full-time workers within 8 km from the workers' place of work; the magnitude of the increase falls sharply thereafter: 0.52% within 8 to 40 km, 0.84% within 40 to 80 km, and 0.20% within 80 to 160 km (11).

Following a similar approach, Di Addario and Patacchini estimate that an increase of 100,000 inhabitants within 4 km raises wages by 0.1% to 0.2%, but the increase falls sharply thereafter (12). The impact of urban size on wages was found to be significant only up to 12 km, which is less than the average radii of Italian local labor markets and thus suggests that agglomeration economies occur within local labor markets. Evidence for Britain by Melo and Graham shows that an increase of 100,000 jobs within 5 km from a workplace raises hourly wages by approximately 1.19%, but the effect falls sharply to 0.38% within 5 to 10 km and 0.15% within 10 to 20 km (13).

Very few papers estimate the decay gradient of agglomeration effects (14–16). Evidence for Britain is obtained by Rice et al. (14) and Graham et al. (16). Rice et al. estimate that the rate of decay of agglomeration effects with increasing driving time is about 1.4, which means that moving the population of working age 30 (60) min farther away decreases the impact of economic mass on productivity by about 75% (94%) (14). Graham et al. obtain a decay gradient of about 1.66 (16). They also find that the decay gradient differs across economic sectors; the value is higher for service industries and smaller for manufacturing. Business services and consumer services have a decay gradient of 1.75 and 1.82, respectively, whereas for manufacturing the value is 1.10. This supports a steeper spatial attenuation of agglomeration externalities for the service sectors, which are generally more dependent on urbanization levels.

The findings from these two papers refer to a catch-all type of measure of agglomeration economies (14, 16); in contrast, Amiti and Cameron focus on the effects from supplier and market access in Indonesia and obtain distance decay parameters of 1.79 and 2.81, respectively (15). The findings indicate that only 10% of the benefits of supplier and market access extend beyond 129 and 82 km, respectively.

## POTENTIAL FOR INTERCITY AGGLOMERATION ECONOMIES

In this section additional considerations are provided on whether agglomeration holds as a useful theory in the context of intercity interactions and on the appropriateness of existing U.K. appraisal practice to capture WEIs from improvements to intercity connectivity.

## Transport and Sources of Agglomeration

Both Marshall's sources of agglomeration (1) and Duranton and Puga's mechanisms of agglomeration (2) bear a relation to transport movements. This is patent in the benefits that accrue to firms from transport investments that facilitate (a) the ease of access to a pool of skilled workers made possible through improvements in the efficiency of commuting flows (labor market pooling, matching); (b) the communication and sharing of ideas between firms and workers (knowledge spillovers, learning, and sharing); and (c) the ease of access to customers and suppliers, including specialized professional services (input–output linkages, sharing, and learning). Some of these benefits are essentially concerned with commuting trips (e.g., labor market pooling, sharing, and matching); others are more related to business trips and freight movement (e.g., input–output linkages, sharing, and learning).

In the case of interurban connections, most of the arguments linking transport to agglomeration could hold in the sense that if spatial interactions between economic agents are made more efficient then increasing returns can be expected. There are no obvious characteristics of the sources or mechanisms described in the agglomeration literature that would limit their generation over longer distances. Labor market effects, knowledge spillovers, and worker input–output interactions could each, in theory, be extended through improvements in temporal accessibility over interregional distances.

## Agglomeration Effects of Transport Investments: Existing Department for Transport Approach

The U.K. DfT has requested that agglomeration benefits be assessed as an additional component of transport appraisal (17). The case for additionality is made clear in Venables, who develops a theoretical model that links agglomeration and transport provision (18). The argument is straightforward: transport investment increases the access that firms have to economic mass, which, if agglomeration economies exist, induces a source of increasing returns that is not captured in a standard transport appraisal. Venables goes on to show that an estimate of the agglomeration benefits of transport investment can be attained if the following are known: (a) the change in access to economic mass that will result from making some transport intervention and (b) the amount by which productivity will rise in response to an increase in agglomeration.

Consequently, such calculations are now being undertaken by transport modelers. DfT uses the concept of functional urban regions to provide guidance on identifying agglomeration benefits from transport schemes. Transport schemes that fall inside functional urban regions and are of considerable scale should include an economic impact report, which quantifies the impacts of the scheme in the local economy (for more details see the DfT *Transport Appraisal Guidance* Unit 3.5.14).

The calculation of the change in productivity ( $y$ ) associated with an increase in agglomeration ( $A$ ) is as follows:

$$\Delta y = \left[ \left( \frac{A_a}{A_b} \right)^{\epsilon_A^y} - 1 \right] y_b \quad (3)$$

where the subscripts  $b$  and  $a$  identify the period before and after the transport intervention, respectively, and  $\epsilon_A^y$  is the productivity elasticity of agglomeration.

The current appraisal of WEIs is essentially concerned with external productivity benefits that can be experienced within urban areas, although it is recognized that the geographic scope over which these benefits are available can diminish, as shown by Graham et al. (16). The parameters used to allow for the spatial decay of productivity benefits are based on distance, not travel times or generalized costs. The distance decay parameters used in the DfT calculations were derived for the appraisal of intraurban investments and ignore the influence of agglomeration effects beyond a 75-km radius. As a result, the methodology may not be appropriate to assess the agglomeration impacts from intercity connectivity.

Moreover, the DfT approach does not distinguish between the effects of transport investments on the sources of agglomeration. Instead, it assumes that a change in access to economic mass achieved through a transport improvement is essentially the same as a proportional shift in the sources of agglomeration economies that are captured within estimated agglomeration elasticities. There is no intermediate stage that attempts to understand the relationship between transport accessibility and the sources of agglomeration. The assumption that transport investments will affect the microfoundations of agglomeration across the board has no empirical basis and is used only to provide an approximation for appraisal purposes. The extent to which the existing approach provides a reasonable approximation is unknown, but it is hard to imagine it will prove consistently reliable across a diverse range of investments.

The extent to which agglomeration economies could be generated through interregional movements can be investigated further by looking at the existing nature of such movements and how they might respond to changes in travel times. The next section provides some empirical evidence on the nature of commuting and business trips for Britain.

## EMPIRICAL EVIDENCE CONCERNING LONG-DISTANCE TRAVEL

In this section gravity models are estimated to provide evidence on the spatial decay of commuting trips and business trips between and within travel-to-work areas (TTWAs). TTWAs are the best available approximations of self-contained labor markets in Britain. They are defined as regions where the proportion of people who live (work) in the area is at least 75% of the total number of people who work (live) in the area. There are nearly 300 TTWAs. Assuming that their geography can be approximated by a circle, the mean (median) area is 772.6 (600.5) km<sup>2</sup>. A current list of TTWA names and codes is available elsewhere (19).

To account for the association between the sources of agglomeration economies and transport movements, distance decay coefficients for commuting trips and business trips are estimated separately. Commuting movements are expected to be connected with labor market pooling externalities (reflecting a more productive matching between job and worker skills), while business movements are expected to be related to input–output linkages (ease of access to customers and suppliers, including specialized professional services) and also knowledge spillovers (learning and sharing of knowledge and practices between businesses).

## Spatial Scope of Commuting Interactions

This section considers the spatial distribution of commuting flows. A considerable volume of commuting beyond TTWA boundaries

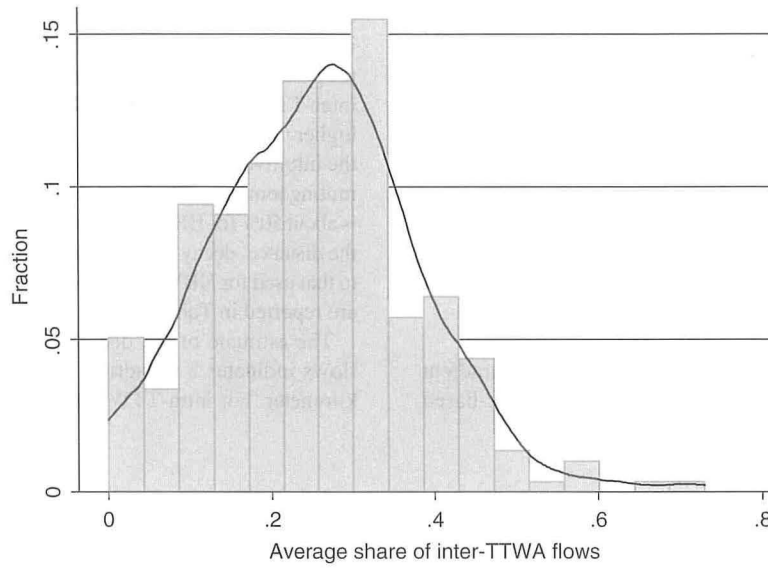


FIGURE 2 Proportion of inter-TTWA HBW flows.

indicates that labor market pooling effects can extend across labor markets. To get an idea about the size of inter-TTWA interactions, the proportion of inter-TTWA commuting for each TTWA in Great Britain was computed with commuting flow data from the national transport model.

Figure 2 shows the histogram of the proportion of inter-TTWA commuting. The values on the x-axis are the proportions of inter-TTWA commuting, and the values on the y-axis are the respective relative TTWA frequencies. The proportion of inter-TTWA home-based work (HBW) trips differs across TTWAs and can reach considerably high shares. The mean value for the proportion of outside TTWA commuting is about 25%, which is just about the same value as the 25% threshold used in the creation of the borders of the TTWAs. But values above this threshold can be as high as 70%. This suggests that there is considerable commuting between different TTWAs although the extent differs across the TTWAs.

To estimate the distance-decay gradient of commuting trips, gravity models are fit with data from the national transport model for HBW movements at the PASS 3 zone level (PASS 3 zones are aggregations of census area statistics wards in 9,998 zones in England, Scotland, and Wales). The standard gravity model for commuting trips can be written as follows:

$$F_{ij} = M_i M_j \exp(-d_{ij}^{-\alpha}) \quad i, j = 1, \dots, N \quad (4)$$

where

- $F_{ij}$  = number of commuting trips between origin  $i$  and destination  $j$ ,
- $M_i$  = mass of the origin consisting of the population at each origin,
- $M_j$  = mass of the destination consisting of the employment at each destination, and
- $d_{ij}$  = Euclidean distance representing spatial distance between each pair of zones.

Equation 4 can be generalized to allow for the presence of origin and destination scale factors that measure the relationship between commuting flows and the respective sizes of origin  $i$  and destination  $j$  (20, p. 22):

$$F_{ij} = c M_i^{\beta_1} M_j^{\beta_2} \exp(-d_{ij}^{-\alpha}) \quad i, j = 1, \dots, N \quad (5)$$

where  $c$  is a constant, and  $\beta_1$  and  $\beta_2$  are parameters that determine the relationship between commuting flows and the size of origins and destinations, respectively.

It is common practice in the empirical literature to apply logarithms to both sides of the gravity equation and then use ordinary least squares to estimate the model. The general model to be estimated is as follows:

$$\log F_{ij} = c + \beta_1 \log M_i + \beta_2 \log M_j + \alpha d_{ij} + D_i + D_j + e_{ij} \quad (6)$$

To control for potential heterogeneity across origins and destinations, a set of control variables for origins ( $D_i$ ) and destinations ( $D_j$ ) based on the DfT rural-urban classification is included. The residual term ( $e_{ij}$ ) is assumed independent and identically distributed according to a normal distribution.

The commuting gravity models for intra- and inter-TTWA flows are estimated separately (Table 1). Using data on long-distance flows only, the decay gradient suggests a reduction in commuting flows of 3.9% for one additional kilometer. Commuting trips inside the same labor market that increase the journey to work by 1 km lead to a fall in flows of 1.7%.

TABLE 1 Gravity Models for Commuting Flows

Statistic	Inter-HBW	Intra-HBW
Log of population in origin	2.1422*	0.9706*
Log of employment in destination	1.7597*	1.4179*
Distance (km)	-0.0394*	-0.0168*
Observations	1,659,416	442,194
Coefficient of determination ( $R^2$ )	.13	.21
Adjusted $R^2$	.13	.21
Control		
Origin	1,652.24*	3,201.7*
Destination	6,595.43*	2,682.79*

\*Significant at .01.

This reduction suggests that commuters may be less sensitive to increases in commute length inside labor markets than to increases in the length of commutes that cross different labor markets. Intuitively, this makes sense since urban transport services inside labor markets are more frequent and better coordinated than between labor markets. The estimates for the mass of origins and destinations indicate positive impacts on commuting flows: the larger the origin (destination) the more commuting flows between the two zones.

### Spatial Scope of Business Interactions

This section presents the results for the distance-decay gradient of home-based employer business (HBEB) and non-home-based

employer business (NHBEB) trips. To get an idea about the size of inter-TTWA interactions for these trips the authors computed the proportion of inter-TTWA flows for each TTWA. The proportion of inter-TTWA HBEB and NHBEB flows (Figure 3) is considerably higher than the proportion of HBW commuting flows. This reflects the intuitive fact that business journeys tend to be longer than commuting journeys. The mean value for the proportion of inter-TTWA is about 50% for HBEB flows and 45% for NHBEB flows. To estimate the distance-decay gradient of business trips a gravity model similar to that used for HBW flows was fit as given in Equation 6. The results are reported in Table 2.

The estimate of the distance gradient for long-distance HBEB flows indicates a reduction in flows of 0.4% for one additional kilometer. For intra-TTWA HBEB trips an increase of 1 km leads

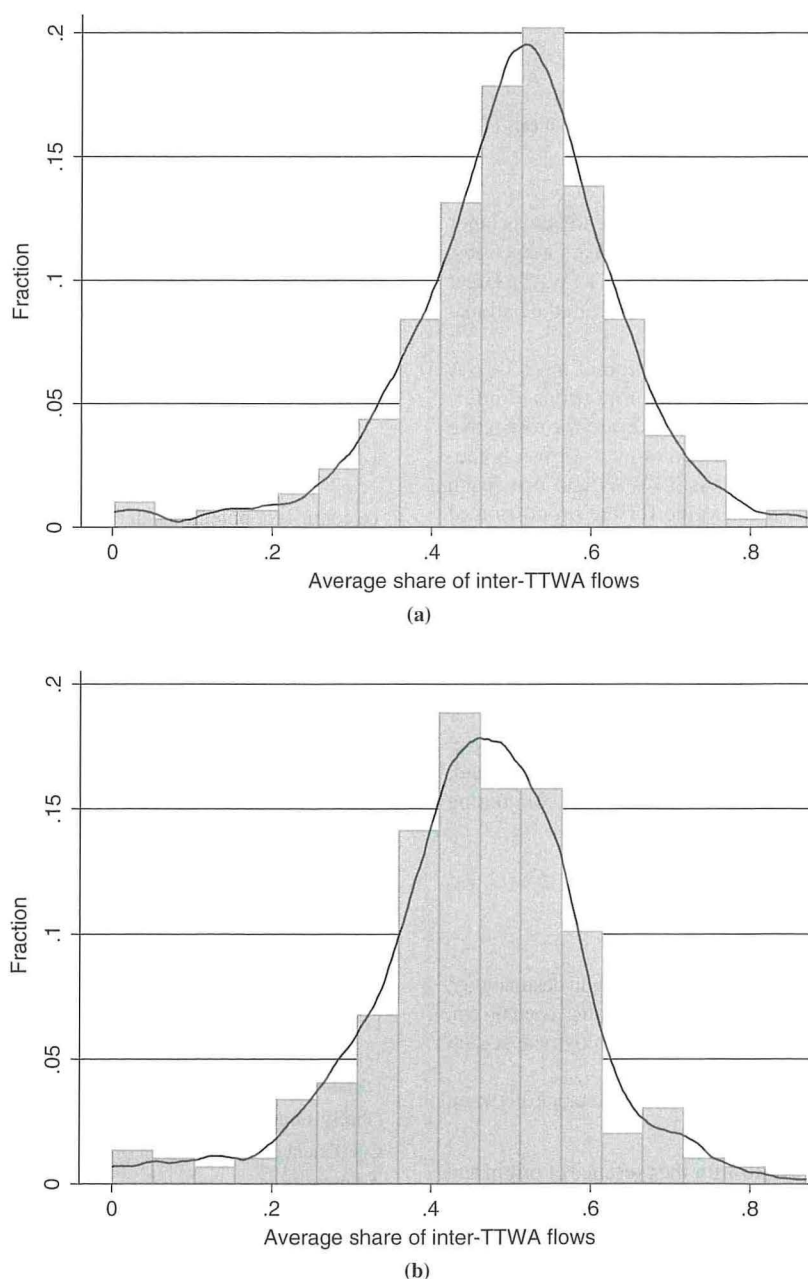


FIGURE 3 Proportion of inter-TTWA flows for (a) HBEB and (b) NHBEB.

TABLE 2 Gravity Models for HBEB and NHBEB Flows

Statistic	Inter-HBEB	Intra-HBEB	Inter-NHBEB	Intra-NHBEB
Log of population in origin	2.0425*	0.9772*	1.3225*	0.3686*
Log of employment in destination	3.9394*	3.3708*	3.3374*	2.3092*
Distance (km)	-0.0040*	-0.0055*	-0.0035*	-0.0053*
Observations	4,055,990	542,932	4,465,602	601,679
Coefficient of determination ( $R^2$ )	.19	.19	.16	.09
Adjusted $R^2$	.19	.19	.16	.09
Control				
Origin	13,198.36*	1,723.54*	7,158.03*	235.04*
Destination	4,784.37*	2,539.50*	9,620.28*	1,806.71*

to a fall in flows of 0.55%. The results for the distance–decay gradient of NHBEB flows are very similar to the values obtained for HBEB flows. The estimate of the distance gradient for long-distance flows suggests a reduction in flows of 0.35% for one additional kilometer. For intra-TTWA flows, the estimate of the decay gradient indicates that an increase of 1 km in distance leads to a fall of 0.53% in flows. The distance–decay gradient of commuting and business movements can be used to infer potential effects of intercity transport investments such as high-speed rail, and ultimately, about the potential order of magnitude of agglomeration benefits. The methodology and calculation of such effects are considered next.

## POTENTIAL ORDER OF MAGNITUDE OF AGGLOMERATION BENEFITS

The calculations presented here, although based on some arbitrary values, offer an approach for the measurement of long-distance agglomeration benefits that could usefully be developed. In particular, this approach addresses the issue of distance versus temporal decay, which is particularly important in the context of long-distance and high-speed investments. The current DfT approach to the assessment of agglomeration benefits allows agglomeration effects to decay spatially, but the decay effect is assumed constant and is relevant only with respect to distance. This approach uses the evidence presented in the fourth section, along with some assumed parameter values (taken from published transport statistics), to allow constancy of decay with respect to equal average travel times but variability with respect to distance. In other words, two equal distances have different decay profiles if travel times vary, but the decay is always the same for equal travel times. This causes the distance decay parameter to change after a transport investment.

The existing DfT approach makes no mention of sources but instead looks at how effective densities change in aggregate and assumes that such changes translate directly into shifts in agglomeration. Another potentially beneficial aspect of the present approach lies in the use of decay estimates for different trip purposes rather than aggregate density changes. This approach could be developed to improve the understanding of the links between transport movements and sources of agglomeration.

## Methodology

The key assumption underpinning this approach is that travel times, not distance, are the dominant factor underlying perceptions of the

attenuation factor for any zone. In developing this methodology it is assumed that the spatial economy is initially at an equilibrium state, in which speeds and flows are consistent.

To represent the relationship between interactions and travel time, the following expression is used:

$$\frac{\partial \log F}{\partial t} = \frac{\partial \log F}{\partial d} \cdot \frac{\partial d}{\partial t} = \alpha \cdot v \quad (7)$$

which is the semi-log derivative of flows ( $F$ ) with respect to travel time ( $t$ ) and is obtained as the product of the distance–decay gradient ( $\alpha$ ) from the gravity models estimated in the fourth section and the (assumed) average speed ( $v$ ) in the system. It measures the proportionate change in interactions given a unit change in travel times. Assuming that this relationship is constant, an increase in average speed is associated with a reduction in the distance–decay gradient, reflecting the fact that with the higher speed distance becomes less of an obstacle to interaction.

The potential magnitude of the agglomeration benefits associated with improvements in travel time resulting from a new high-speed rail connection ( $WEI_{HS}$ ) is obtained as follows:

$$WEI_{HS} = \left( \left[ \sum_{K=1}^3 w_K \cdot \frac{\Delta ED}{ED} \right] \cdot \epsilon_{ED}^v \right) \cdot GDP \quad (8)$$

where

$w_K$  = proportion of long-distance (inter-TTWA) flows for purpose  $K$ , with  $K = 1$  (HBW), 2 (HBEB), and 3 (NHBEB);

ED = measure of effective density (defined in Equation 9);

$\epsilon_{ED}^v$  = output elasticity with respect to ED (16); and

GDP = gross domestic product.

Effective density (ED) is defined as

$$ED = \sum_i ED_i = \sum_i \sum_j \left( \frac{A_j}{\exp(\alpha^{SD} \cdot d_{ij})} + \frac{A_j}{\exp(\alpha^{LD} \cdot d_{ij})} \right) \quad i \neq j \quad (9)$$

where  $A_j$  is a measure of the level of opportunities at destination zone  $j$ , and SD and LD are short distance and long distance, respectively.  $A_j$  is measured by population for HBW flows, population and employment for HBEB flows, and employment for NHBEB flows.

The change in the effective density resulting from an increase in the average speed of long-distance flows after a transport investment is incorporated through a new (smaller) distance–decay gradient:

$$\begin{aligned} \frac{\Delta ED}{ED} &= \frac{ED^1 - ED^0}{ED^0} \\ &= \sum_i^N \sum_j^M \left( \frac{A_j}{\exp(\alpha^{SD} \cdot d_{ij})} + \frac{A_j}{\exp(\alpha_i^{LD} \cdot d_{ij})} \right) \\ &\quad - \sum_i^N \sum_j^M \left( \frac{A_j}{\exp(\alpha^{SD} \cdot d_{ij})} + \frac{A_j}{\exp(\alpha_0^{LD} \cdot d_{ij})} \right) \\ &= \frac{\sum_i^N \sum_j^M \left( \frac{A_j}{\exp(\alpha^{SD} \cdot d_{ij})} + \frac{A_j}{\exp(\alpha_i^{LD} \cdot d_{ij})} \right)}{\sum_i^N \sum_j^M \left( \frac{A_j}{\exp(\alpha^{SD} \cdot d_{ij})} + \frac{A_j}{\exp(\alpha_0^{LD} \cdot d_{ij})} \right)} \end{aligned} \quad (10)$$

where the after-transport-improvement distance–decay gradient for long-distance journeys is

$$\alpha_i^{LD} = \left( \frac{\partial \log F}{\partial d} \cdot \frac{\partial d}{\partial t} \right) \cdot (v_i^{LD})^{-1} \quad (11)$$

and the after-transport-improvement average speed is

$$v_i^{LD} = k \cdot v_0^{LD} \quad k > 0 \quad (12)$$

where  $k$  is a positive constant that reflects the increase in average speed. Thus, in this approach, the distance–decay parameter is assumed nonconstant and is revised after a transport intervention changes the average speed in the transport network. The intuition here is that improvements in travel times (resulting from faster journeys) will reduce the extent to which travelers perceive distance as an obstacle to interaction.

### Order of Magnitude of Agglomeration Benefits

Some illustrative calculations are provided to suggest the orders of magnitude of agglomeration benefits that could arise from investments affecting long-distance movements. These are illustrative only, and should not be taken as estimates of the likely agglomeration effects of high-speed rail.

The steps and assumptions used to quantify the potential magnitude of agglomeration benefits are explained below. The assumptions used are based on information for the proposals of a new high-speed railway line linking London to the West Midlands, Northern England, and Scotland (21, 22). This proposed high-speed rail line is known as High-Speed 2 (HS2); HS1 is the high-speed rail line linking London to continental Europe through the Channel Tunnel Rail Link.

According to the demand model analysis of the proposed HS2 project (21, 22), average travel times are expected to improve between 25% and 50%, depending on the particular rail line corridor and route design. Two scenarios for travel time improvements caused by HS2 are considered: Scenario 1 assumes an improvement of 25% (lower bound), and Scenario 2 assumes an improvement of 50% (upper bound).

In addition to the improvement to travel times, the potential effect of high-speed rail on rail's market share must be considered. According to the National Travel Survey, rail accounts for about 12% of all long-distance trips. The improvements in rail travel times brought about by a new high-speed rail line are expected to increase rail's market share in long-distance trips, both through mode shift from alternative modes (e.g., air, road) and new trip generation. Modal shift analysis for the proposed HS2 project estimates the sources of the additional high-speed and conventional rail passengers, averaged across the different corridors, to be 11% from road, 20% from air, and 60% from new trips (22, p. 92). Modal shift analysis further estimates that 66% of the new high-speed rail passengers will come from classic rail, 8% from road, 6% from air, and 23% from new trips (22, p. 93). On the basis of these values, and without any further information, it is assumed that high-speed trips account for half of all long-distance rail trips.

The attractiveness of high-speed rail will also be determined by factors other than travel times. Differences in fares, service frequency, and comfort level between conventional and high-speed rail are also expected to influence mode choice. Although no explicit account of such differences is taken, it is expected that they are reflected in the market share of high-speed rail.

Tables 3 and 4 summarize the values obtained for each of the terms in Equation 8 and the assumptions described above. Table 3 provides an order of magnitude for agglomeration benefits according to two scenarios of travel time improvements: (a) Scenario 1 with an improvement of 25% and (b) Scenario 2 with an improvement of 50%. In Table 4 each of the two scenarios is combined with three hypotheses for the growth of rail's market share of long-distance

**TABLE 3 Agglomeration Benefits from Improvements in Long-Distance Travel Times**

Statistic	HBW	HBEB	NHBEB
Distance decay ( $\alpha^{LD}$ ) before HS intervention	−0.0394	−0.0040	−0.0035
Initial average travel time (all flows) <sup>a</sup> (min)	27	39	
Share of travel purpose $K$ in LD flows ( $w_K$ ) <sup>b</sup>	0.61	0.17	0.22
Scenario 1. 25% reduction in travel times			
Distance decay ( $\alpha^{LD}$ ) after HS intervention	−0.0315	−0.0032	−0.0028
Change in ED ( $\Delta ED/ED$ )	0.11	0.10	0.09
Scenario 2. 50% reduction in travel times			
Distance decay ( $\alpha^{LD}$ ) after HS intervention	−0.026	−0.0027	−0.0023
Change in ED ( $\Delta ED/ED$ )	0.21	0.17	0.15

NOTE: HS = high speed.

<sup>a</sup>Transport Statistics Great Britain and Royal Automobile Club Foundation and the British Chambers of Commerce.

<sup>b</sup>Based on data from national transport model (as used in gravity models).

**TABLE 4** Hypotheses for Rail Market Share and Agglomeration Benefits

Statistic	Share or Benefit
Share of rail in LD trips <sup>a</sup> (%)	12
Share of HS rail trips in all LD rail trips (%)	50
Share of rail in LD trips with HS (%)	
Pessimistic scenario	12
Moderate scenario	25
Optimistic scenario	50
Output elasticity with respect to effective density <sup>b</sup>	0.04
GDP (£ million) <sup>c</sup>	1,446,113
Scenario 1. WEIs from HS rail (£ million,% GDP)	
Pessimistic scenario	346.88, 0.02
Moderate scenario	722.67, 0.05
Optimistic scenario	1,445.33, 0.10
Scenario 2. WEIs from HS rail (£ million,% GDP)	
Pessimistic scenario	661.59, 0.05
Moderate scenario	1,378.32, 0.10
Optimistic scenario	2,756.64, 0.19

<sup>a</sup>National Travel Survey, Table NTS0317: share of rail trips in total long-distance trips within Britain.

<sup>b</sup>Graham et al. (16).

<sup>c</sup>Office for National Statistics for 2008 (*United Kingdom National Accounts 2009: The Blue Book*).

trips as a result of the opening of a high-speed rail line: (a) a pessimistic hypothesis, assuming no change in rail's market share; (b) a moderate hypothesis, assuming an increase in rail's market share from 12% to 25%; and (c) an optimistic hypothesis, assuming an increase in rail's market share from 12% to 50%. These values should be viewed as averages across the various HS2 corridors: higher rail market shares are expected for shorter corridors (e.g., London to West Midlands and Northern England), and smaller shares are anticipated for longer corridors (London to Scotland).

Under Scenario 1 (improvement to travel times of 25%) and assuming mode shares as described above, agglomeration benefits of £346.8 million (pessimistic hypothesis), £722.67 million (moderate hypothesis), and £1,445.33 million (optimistic hypothesis) for rail market share are estimated (£1 = \$1.60 in 2011 U.S. dollars). In relative terms, these benefits correspond to 0.02%, 0.05%, and 0.10% of the GDP, respectively.

Under Scenario 2 (improvement to travel times of 50%), agglomeration benefits of £661.59 million (pessimistic hypothesis), £1,378.32 million (moderate hypothesis), and £2,756.64 million (optimistic hypothesis) for rail market share are estimated. In relative terms, these benefits correspond to 0.05%, 0.10%, and 0.19% of the GDP, respectively.

Taking the moderate hypothesis for both scenarios and the official estimated value of HS2's benefits (21), the magnitude of the agglomeration benefits is about 2.5% and 4.8% of the present value of the conventional transport benefits for Scenario 1 and Scenario 2, respectively.

These calculations can provide a lower and upper bound for the potential order of magnitude of agglomeration benefits from the new high-speed rail line. The worst-case scenario predicts a benefit of £346.8 million, equivalent to 0.02% of the GDP of the United Kingdom. The best-case scenario predicts a benefit of £2,756.64 million, equivalent to 0.19% of the GDP of the United Kingdom. The main conclusion is that even under a very optimistic scenario for the improvement in long-distance travel times and the market share of

classic and high-speed rail trips, the potential order of magnitude of the agglomeration benefits is small.

## CONCLUSIONS

This paper has considered the scope for agglomeration-based WEIs in the context of high-speed rail investment. It has provided an overview of the theoretical mechanisms that drive agglomeration-based WEIs, generated new empirical evidence on long-distance travel, and provided an indicative assessment of the potential order of magnitude of agglomeration benefits from long-distance transport improvements.

Theory suggests that the main sources of agglomeration externalities are caused by labor market pooling, knowledge spillovers, and efficiency in input-output linkages. There are good reasons to believe that transport investments can affect the microfoundations of agglomeration. Furthermore, there are no obvious characteristics of the sources or mechanisms of agglomeration discussed in the literature that would limit their generation over longer distances. Thus, most of the arguments linking transport to agglomeration could hold in long-distance cases in the sense that if spatial interactions between economic agents are made more efficient, then increasing returns could be expected.

Evidence on the spatial distribution of commuting and business flows shows a considerable amount of long-distance interactions between TTWAs, especially for business interactions. This suggests that transport improvements to long-distance journeys such as high-speed rail could have an important effect on the level of connectivity between firms and workers.

With estimates of the distance decay of commuting and business flows obtained from the estimation of the gravity models, it is possible to infer changes in flows resulting from improvements in travel times. To estimate potential agglomeration benefits from improvements to long-distance travel, an approach is adopted that assumes constancy of trip decay with respect to equal average travel times but variability with respect to distance. The intuition is that improvement in travel times will reduce the extent to which travelers perceive distance as an obstacle to interaction.

The potential order of magnitude of agglomeration benefits is estimated to be between a lower bound of £346.8 million (0.02% of U.K.'s GDP) and an upper bound of £2,756.64 million (0.19% of U.K.'s GDP). Consequently, even in the best-case scenario for improvement in long-distance travel times and market share of classic and high-speed rail, the potential order of magnitude of the agglomeration benefits is expected to be small.

The methodology proposed in this paper considers only the potential benefits within Britain's domestic market, but benefits could also arise from improved connections to continental Europe (e.g., Paris, Brussels, Belgium; and Amsterdam, the Netherlands) by linking HS2 to HS1.

## REFERENCES

1. Marshall, A. *Principles of Economics*, 8th ed. Macmillan, London, 1920. Library of Economics and Liberty. <http://www.econlib.org/library/Marshall/marP.html>.
2. Duranton, G., and D. Puga. Micro-Foundations of Urban Agglomeration Economies. In *Handbook of Regional and Urban Economics* (J. V. Henderson and J. F. Thisse, eds.), Elsevier, Amsterdam, Netherlands, 2004.

3. Graham, D. J. Agglomeration, Productivity and Transport Investment. *Journal of Transport Economics and Policy*, Vol. 41, No. 3, 2007, pp. 317–343.
4. Graham, D. J. Variable Returns to Urbanization and the Effect of Road Traffic Congestion. *Journal of Urban Economics*, Vol. 62, No. 1, 2007, pp. 103–120.
5. Rosenthal, S. S., and W. C. Strange. Evidence on the Nature and Sources of Agglomeration Economies. In *Handbook of Urban and Regional Economics, Volume 4* (J. V. Henderson and J. F. Thisse, eds.), Elsevier, Amsterdam, Netherlands, 2004.
6. Melo, P. C., D. J. Graham, and R. B. Noland. A Meta-Analysis of Estimates of Urban Agglomeration Economies. *Regional Science and Urban Economics*, Vol. 39, No. 3, 2009, pp. 332–342.
7. Graham, D. J. *Wider Economic Benefits of Transport Improvements: Link Between Agglomeration and Productivity*. Stage 2 report. Imperial College of Science, Technology, and Medicine, London, 2006.
8. Ahmed, M. E. G., and D. M. Levinson. *Access to Destinations: Development of Accessibility Measures*. Minnesota Department of Transportation, Saint Paul, 2006.
9. Vickerman, R., K. Spiekermann, and M. Wegener. Accessibility and Economic Development in Europe. *Regional Studies*, Vol. 33, No. 1, 1999, pp. 1–15.
10. Combes, P.-P., and M. Lafourcade. Transport Costs: Measures, Determinants, and Regional Policy Implications for France. *Journal of Economic Geography*, Vol. 5, No. 3, 2005, pp. 319–349.
11. Rosenthal, S. S., and W. C. Strange. The Attenuation of Human Capital Spillovers. *Journal of Urban Economics*, Vol. 64, No. 2, 2008, pp. 373–389.
12. Di Addario, S., and E. Patacchini. Wages and the City: Evidence from Italy. *Labour Economics*, Vol. 15, No. 5, 2008, pp. 1040–1061.
13. Melo, P. C., and D. J. Graham. *Agglomeration Economies and Labour Productivity: Evidence from Longitudinal Worker Data for GB's Travel-to-Work Areas*. Spatial Economics Research Center Discussion Paper 31, London School of Economics, London, 2009.
14. Rice, P., A. J. Venables, and E. Patacchini. Spatial Determinants of Productivity: Analysis for the Regions of Great Britain. *Regional Science and Urban Economics*, Vol. 36, No. 6, 2006, pp. 727–752.
15. Amiti, M., and L. Cameron. Economic Geography and Wages. *Review of Economics and Statistics*, Vol. 89, No. 1, 2007, pp. 15–29.
16. Graham, D. J., S. Gibbons, and R. Martin. Transport Investments and the Distance Decay of Agglomeration Benefits. Working paper. Imperial College of Science, Technology, and Medicine, London, 2009.
17. *Transport, Wider Economic Benefits and Impacts on GDP*. Department for Transport, Her Majesty's Stationery Office, London, 2005.
18. Venables, T. Evaluating Urban Transport Improvements: Cost–Benefit Analysis in the Presence of Agglomeration and Income Taxation. *Journal of Transport Economics and Policy*, Vol. 41, 2007, pp. 173–188.
19. Office for National Statistics. [http://www.statistics.gov.uk/geography/geographic\\_area\\_listings/other.asp](http://www.statistics.gov.uk/geography/geographic_area_listings/other.asp).
20. Fotheringham, A. S., and M. E. O'Kelly. *Spatial Interaction Models: Formulations and Applications*. Studies in Operational Regional Science, Vol. 5, Kluwer Academic Publishers, Dordrecht, Netherlands, 1989.
21. *High Speed Rail*. Command paper. Department for Transport, London, 2010.
22. *HS2, High-Speed Rail London to the West Midlands and Beyond: A Report to Government by High-Speed Two Limited*. Department for Transport, London, 2010.

---

*The Intercity Passenger Rail Committee peer-reviewed this paper.*

# Economic Impacts of Intercity Passenger Rail Service

## Evidence from Passenger Surveys

Benjamin R. Sperry and Curtis A. Morgan

Recent policy and funding developments have stimulated interest in high-speed and intercity passenger rail. Given the present economic climate of the country, particular interest lies in the impacts of new or expanded passenger rail services on local or regional economies. Using evidence from rail passenger surveys, this paper examines the economic impacts of existing state-supported passenger rail service in a growing intercity corridor of the south central United States. Economic impacts were measured by analyzing passengers' self-reported expenditures on selected items during their rail trip. The results of the analysis showed an estimated spending impact on local communities of more than \$18 million in total, with more than \$1.3 million of local sales tax generated by rail passengers in the study corridor. If the service were discontinued, an estimated 30% of these impacts would be lost. The findings of this analysis inform passenger rail planning efforts by providing better understanding of the financial considerations and also assist rail planners in developing local support for new rail projects.

Recent developments have stimulated interest in high-speed and intercity passenger rail throughout the United States. The October 2008 passage of Public Law 110-432, the Passenger Rail Improvement and Investment Act of 2008 (PRIIA), established new and expanded policy, funding, and regulatory roles for the federal government in the development of high-speed intercity passenger rail service (1). Funding of high-speed and intercity passenger rail capital investment grant programs established in PRIIA, which included substantial requirements for state and local matching funds for federal grants, was made available through \$8 billion of appropriations from the American Recovery and Reinvestment Act of 2009 (2). Collectively, these events have generated a significant amount of interest in high-speed intercity passenger rail among planners, policy makers, and the public at large.

The new policy and regulatory direction for intercity passenger rail will ultimately lead to new public investment in intercity corridors where passenger rail could significantly improve mobility, safety, air quality, energy consumption, and the livability of major urban areas. However, because of the current economic climate of the country, many state budgets are facing deficits and are having difficulty meeting their current obligations. Furthermore, while infrastructure investments such as the ones necessary to renew the U.S. passenger rail

network are attractive policy directions for short-term job creation and long-term mobility benefits, tough decisions on funding new intercity passenger rail routes or upgrades to existing services will be necessary. Consequently, there exists a need to identify and understand the underlying economic impacts of investing in high-speed or other intercity passenger rail systems. This paper contributes to a growing body of literature on the economic impacts of existing intercity passenger rail by describing the findings of a case study of the economic impacts associated with rail passenger spending in communities served by intercity passenger rail corridor service.

### LITERATURE REVIEW

Measuring and understanding the total economic impacts of transportation investments is one of the most important pieces of the project development puzzle. Understanding these impacts can better inform policy makers by providing an accurate picture of how a proposed project will affect public revenue sources, job creation, or direct spending related to project activities. Clearly articulating these impacts to the public is critical for obtaining project buy-in from local residents and elected officials in the communities where projects are proposed. The economic impacts of a transportation project include the following (3):

- Direct user benefits. Traveler cost savings as a result of improved service levels, ease of access, comfort, or safety resulting from a transportation project;
- Direct economic benefits. Monetary benefits for some facility users and nonusers in the affected area of a transportation project as a result of reduced logistics costs, increased business activity, or increased accessibility to new regions;
- Indirect or induced impacts. Spin-off effects of new investment along the supply chain, such as suppliers of goods or services to industries directly involved with new transportation projects, and locational shifts in population and businesses; and
- Impacts of construction and maintenance spending. The short-term impacts of the construction of a transportation facility and the long-term impacts of maintenance or operations of a transportation facility.

The measurement of economic impacts related to transportation projects can be accomplished through several approaches (3). Simulation models, such as input-output models, use information on technologies and local trade to estimate economic impacts by applying known multipliers to these inputs. The Bureau of Economic Analysis regional input-output modeling system (RIMS-II) is an example of

Texas Transportation Institute, Texas A&M University System, 3135 TAMU, College Station, Texas 77845-3135. Corresponding author: C. A. Morgan, c-morgan@ttimail.tamu.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 25–30.  
DOI: 10.3141/2261-03

such a model. Often, results from a larger transportation system model are applied to an economic simulation model to calculate the full impact of a new transportation project. For transportation projects with a significant expenditure of public resources (such as what is necessary for large-scale passenger rail projects), a great deal of emphasis is placed on the favorable economic impacts of the proposed project. Feasibility studies of proposed high-speed rail routes in the Midwest (4, 5), Florida (6), and California (7, 8) have identified a number of potential economic impacts of new high-speed rail investment in terms of job creation and user benefits.

In general, economic impact studies of proposed high-speed passenger rail routes use an integrated transportation system and economic simulation modeling approach. A second approach to measuring economic impacts, the direct measurement technique, does not use simulation but instead relies on primary (surveys of travelers or businesses) or secondary (employment or income changes) data sources for the identification of impacts (3). However, for proposed routes, direct measurement techniques are not an option since the services are not yet in operation. For economic impact studies of routes that are currently in operation, the direct measurement approach can be used to identify the economic impact of current, conventional (i.e., 79 mph) intercity rail service. Data sources used in the direct measurement approach include, for example, interviews of rail passengers or local stakeholders in communities served by rail lines. While the total project impacts (jobs and system user benefits) are important for policy formulation and transportation planning decisions, direct impacts to communities served by current or proposed services are important tools for obtaining local stakeholder support for passenger rail projects.

A 2005 study examining the economic benefits of the *Amtrak Downeaster* intercity passenger rail route (Boston, Massachusetts, to Portland, Maine) used an onboard passenger survey to identify total passenger spending on lodging, food and beverage, entertainment, retail, and local transportation purchases. Using these responses, the study estimated that visitors to Maine and New Hampshire who rode the *Downeaster* rail service contributed \$3.5 million in direct spending to the communities served by the route. Additionally, the study used findings from site interviews of businesses, real estate developers, and other stakeholders in communities served by the route to identify the new economic development related to the rail service, which was estimated to be \$4.5 million annually to these communities (9).

In terms of data and findings obtained with the direct measurement approach, the *Downeaster* study appears to be the most comprehensive study of the economic impacts of current intercity passenger rail service available in the literature. The *Downeaster* study is particularly notable in that visitor spending levels were estimated directly from passenger survey data. Alternative approaches are also found in the literature. A study of the community impacts of the three Amtrak routes in Michigan used onboard passenger survey data to identify factors such as trip purpose, travel duration, and travel mode used by the passenger to access the rail boarding station and then applied assumed expenditures for each area to estimate total economic impacts (10).

A fast-growing segment of Amtrak's service is its corridor service segment, which focuses on state-supported routes between 100 and 500 mi (161 and 805 km) in length. Examples of this type of service include the *Hiawatha* service between Chicago, Illinois, and Milwaukee, Wisconsin, as well as heavily used routes in California and the Pacific Northwest. The service times and frequencies available for such short- to medium-distance routes make them popular to a wide range of riders, and a number of measurable benefits have been identified in terms of providing intercity mobility in these corridors (11). A lesser-known but growing route is the *Heartland Flyer* service, a low-

density corridor service that currently operates between Oklahoma City, Oklahoma, and Fort Worth, Texas. This route is the focus of the case study described in this paper. The route's modest schedule (single daily round-trip) and service area in the south central United States, outside the traditional areas for intercity passenger rail, make it a good case study to determine the possible economic impacts that might accrue from the introduction of new intercity passenger rail service in other areas where it may not currently operate. Information on the economic impacts resulting from such a minimal service provides a baseline for estimating the economic impacts of additional service frequencies and travel time improvements that could stem from incremental investments or the introduction of new high-speed service.

## CASE STUDY OVERVIEW

### Description of Case Study Rail Route

The intercity passenger rail route examined in this analysis is the *Heartland Flyer*, which operates along a 206-mi (332-km) route between Oklahoma City, and Fort Worth. Intermediate communities served by the *Heartland Flyer* include Norman, Purcell, Pauls Valley, and Ardmore in Oklahoma and Gainesville in Texas. The *Heartland Flyer* is operated by the National Railroad Passenger Corporation (Amtrak) with financial support provided by a two-state partnership between Oklahoma and Texas. The route makes one round-trip between the corridor endpoints daily, traveling southbound from Oklahoma City to Fort Worth in the morning and returning northbound in the evening (12). More than 657,000 passengers have made trips on the service since its inauguration in June 1999. For the 12-month period ending September 2009 (FY 2009), 73,564 passengers rode on the *Heartland Flyer*, and ridership has grown consistently at approximately 8% annually since 2003. Since 1999, ticket revenues have totaled more than \$12.2 million, and revenue from food and beverage has exceeded \$1.4 million. The intercity corridor over which the route operates is a section of one of three spokes of the federally designated south central high-speed rail corridor (2). Current (FY 2009) economic impacts of the *Heartland Flyer* service on jobs, wages, and direct spending are estimated as follows (author analysis of unpublished data provided by Amtrak):

- **Jobs.** The *Heartland Flyer* provides 10 full-time positions directly involved with daily onboard operations (engineers, conductors, food service) plus an additional six positions for ticketing, maintenance, and operational support in Fort Worth.
- **Wages and benefits.** Estimated annual wages and benefits are \$1,538,000 for the 16 employees associated with the *Heartland Flyer* service.
- **Direct spending.** Amtrak purchases of fuel and supplies related to *Heartland Flyer* operations include fuel purchases in Fort Worth valued at \$562,000 annually and \$803,200 in the state of Oklahoma for unspecified goods and services.

In April 2005, a study prepared for the Oklahoma Department of Transportation examined the economic benefits of the *Heartland Flyer* service to Oklahoma (13). The study reported more than \$11.4 million of direct spending attributed to the service, including earnings of Oklahoma residents employed by the service as well as tax revenues. The indirect effect of the state's investment in the *Heartland Flyer* was reported at more than \$22 million over the 5-year period since the service was initiated.

## Survey Design and Implementation

The two-page survey instrument used in this study contained 21 questions about the passengers' origin and destination, travel modes to and from the rail station, trip purpose, alternative travel mode, trip frequency, reasons for choosing the *Heartland Flyer*, personal spending, and demographics. The data collection procedure was a hand-out-hand-back scheme, with data collection staff passing through the train distributing surveys at the start of each train run and periodically passing back through to collect completed survey forms from the passengers. Data were collected in two rounds in April and July of 2009. Analysis of 120 months of historical ridership data (October 1999 to September 2009) for the *Heartland Flyer* indicates that the ridership during these 2 months represents the average (April) and peak (July) travel demand periods for the route (11). While more rounds of surveying would have been helpful to capture additional seasonal effects on passenger spending (such as spending for shopping during the holiday periods), the two periods surveyed as part of this project were deemed to be sufficiently representative of the ridership as a whole.

During the April data collection, 877 passengers boarded the *Heartland Flyer* during 10 data collection runs, and 435 valid surveys were obtained. In July, the 1,161 passengers who rode the *Heartland Flyer* during seven data collection runs returned 588 valid survey forms. During both rounds, approximately two-thirds of the passengers were eligible to participate in the study. Ineligible passengers included minors and passengers traveling as part of an organized group. For both April and July, the total participation rate was 76% of those eligible. In general, participation rates for intercity passenger rail onboard surveys are high, likely because the passenger is captive on the train for a long enough time to complete the survey (14).

The specific economic impact evaluated in this analysis was the total impact of expenditures generated by *Heartland Flyer* passengers during their rail trip. Direct measurement of passenger spending during the rail trip was accomplished through a single question on the passenger survey. The question asked the passenger to report or estimate how much he or she would spend on four specific items (lodging, meals, shopping, and entertainment) during his or her trip. The question was worded as follows: "How much did you spend (or plan to spend) on lodging, meals, shopping, or entertainment during the entire duration of your trip?" Passengers were provided five options with a range of spending levels to provide as a response: less than \$25, \$25 to \$49, \$50 to \$99, \$100 to \$249, and \$250 or more. Passengers were requested to not include purchases made onboard the train (i.e., food purchases from the cafe car) as these impacts are tracked through other means.

The use of a single question to identify passenger spending in this study was a different approach than that used in the *Downeaster* economic impact study, which requested a more detailed estimate of spending as separate items, but used similar categories (9). The response rate for the passenger spending item was 98% of all valid surveys returned in April (average season) and 92% in July (peak season). The item nonresponse rate for the equivalent questions in the *Downeaster* study was not available. Therefore, the item nonresponse effects of having a single question to encompass all spending categories, rather than a series of questions on spending, could not be examined. One potential issue that was identified with this question was that it was necessary for some travelers to estimate spending levels, as the actual spending had not occurred at the time the survey was administered. However, if such an estimate was required in this situation, it was likely that the actual spending levels were probably greater than estimated, rather than less. Since this error would generally result in a conservative estimate of passenger spending, it was not considered a major issue for the final analysis.

TABLE 1 Percentage of Passengers by Spending Range

Spending Range	Average Season (%)	Peak Season (%)
Less than \$25	16	11
\$25 to \$49	11	12
\$50 to \$99	19	15
\$100 to \$249	28	28
\$250 or more	27	35

NOTE: Columns may not add up to 100% because of rounding.

SOURCE: 2009 *Heartland Flyer* passenger survey (11).

The raw data were entered into a computer spreadsheet program and underwent a thorough quality control process that identified and corrected many common errors found in intercity rail passenger onboard survey data. More details of the data collection and quality control process can be found in the full project report (11).

## ANALYSIS OF PASSENGER SPENDING

The first phase of the spending analysis identified the general spending levels of *Heartland Flyer* passengers as well as trends and patterns in spending behavior observed from key survey variables. Table 1 reports the percentage of passengers reporting each level of spending for the average and peak seasons. Responses indicate that a majority of passengers spent more than \$100 during their rail trip. During the average season, the most frequent spending range reported by passengers was the \$100 to \$249 range (28%), while the most frequent range for the peak season was the "\$250 or more" category (35%). Median spending levels for each season (in Table 1 and subsequent tables) were calculated by interpolating between the endpoints of the spending range that contained the 50th percentile of the cumulative distribution of responses. The median passenger spending levels for each season were calculated to be \$123.91 for the average season and \$166.67 for the peak season. The difference in the distribution of reported spending levels between the average and peak seasons was significant ( $\chi^2 = 12.49$ ,  $p = .0141$ ).

On the basis of seasonality differences identified in Table 1, certain trip-related variables obtained from the onboard passenger survey were suspected to influence passenger spending levels. Specifically, passenger spending related to two trip-related variables, passenger trip purpose and duration of trip away from home, was examined in greater detail. One might expect these variables to reflect seasonal spending differences as longer and more expensive travel may take place during the summer, when more time is available for family travel or vacations. Table 2 reports the median spending levels of *Heartland Flyer* passengers by self-reported trip purpose. Trip purposes of "visiting family or friends" or "leisure, recreation, or

TABLE 2 Median Passenger Spending Level by Trip Purpose

Trip Purpose	Average Season (\$)	Peak Season (\$)
Visiting family or friends	74.29	91.84
Personal, business, or education	86.59	91.58
Leisure, recreation, or shopping	195.62	206.65
Vacation	204.59	250 or more <sup>a</sup>

<sup>a</sup>Median could not be calculated because category was open-ended.

SOURCE: 2009 *Heartland Flyer* passenger survey (11).

shopping” were reported by a majority of passengers (82% during the average season, 77% during the peak season). The trend identified in Table 1 of higher reported spending during the peak season was apparent in the breakdown of spending by trip purpose in Table 2 as well. The difference in spending levels among the four trip purposes was significant during both seasons (average  $\chi^2 = 57.92$ ,  $p < .0001$ ; peak  $\chi^2 = 92.091$ ,  $p < .0001$ ).

Passengers reporting a trip purpose of “visiting family or friends” or “personal, business, or education” spent significantly less during their trips than passengers reporting the other two trip purposes. Trips for “leisure, recreation, or shopping” had a reported a median spending level of approximately \$200 during both seasons. Passengers with “vacation” as a trip purpose reported the highest median spending levels among all the trip purpose categories, spending approximately \$205 during the average season. During the peak season, more than 70% of vacation travelers reported spending in excess of \$250 during their trips. Since this category was open-ended, the median spending level for vacation travelers in the peak season could not be calculated with the available data. Vacation travelers also had the largest difference in spending between the two seasons, spending at least \$50 more during the peak season than the average season, as compared to a difference of less than \$20 for the other three purposes.

Passengers were also asked to report the duration of their trip away from home, given as the number of nights away from home during the trip. Five response categories were provided, including zero nights (i.e., the passenger was making a round trip that day). Table 3 reports the median passenger spending levels by the five trip duration categories. The difference in spending levels among the five trip durations was significant during both seasons (average  $\chi^2 = 51.034$ ,  $p < .0001$ ; peak  $\chi^2 = 50.007$ ,  $p < .0001$ ). Surprisingly, the median passenger spending levels did not increase as the trip duration increased. Rather, the highest spending level was by passengers reporting two nights away from home (\$188.14) during the average season and for passengers reporting three to five nights away from home (\$232.20) during the peak season.

Demographic variables were also examined for spending patterns. No significant differences in spending between men and women were observed for either the average season ( $\chi^2 = 1.269$ ,  $p = .8666$ ) or the peak season ( $\chi^2 = 9.966$ ,  $p = .062$ ). However, differences in spending patterns among passengers of different age groups were significant (average  $\chi^2 = 57.094$ ,  $p < .0001$ ; peak  $\chi^2 = 37.167$ ,  $p < .0112$ ). These differences were reflected primarily among younger travelers, who appeared to spend less during their trips than other age groups. Differences in spending patterns among passengers reporting different annual household income levels were significant (average  $\chi^2 = 75.149$ ,  $p < .0001$ ; peak  $\chi^2 = 63.793$ ,  $p = .0001$ ). The responses indicated that as passenger household income increased, reported spending on the trip increased as well, which was not a surprising finding.

**TABLE 3 Median Passenger Spending Level by Trip Duration**

Trip Duration	Average Season (\$)	Peak Season (\$)
Round-trip today	71.00	94.74
One night	182.86	205.59
Two nights	188.14	177.78
Three to five nights	118.60	232.20
Six or more nights	125.00	178.50

SOURCE: 2009 *Heartland Flyer* passenger survey (11).

## PASSENGER SPENDING IMPACTS

The second phase of the analysis examined the spending data with a policy focus, investigating spending patterns to determine the potential spending impacts of rail passengers at each community served by the *Heartland Flyer* route. To extract community spending impacts from the survey data, a destination station was assigned to each survey record on the basis of the boarding and alighting station and the passenger’s residential zip code. These destination stations were assumed to be the location where the spending occurred during the rail trip. Researchers computed the average passenger spending level at each station per study period by using a midpoint value assigned to each range of spending given as choices on the survey form. Total spending was then calculated by distributing the average spending level for each station across the FY 2009 boarding and alighting totals for each station to either peak or nonpeak travel seasons. An adjustment was made to the passenger activity in Fort Worth to account for passengers transferring to another Amtrak train (the *Texas Eagle*) as the expenditures by these passengers do not remain in the Fort Worth area. With the destination station information, the impacts of passenger spending on *Heartland Flyer* communities were estimated.

### Total Community Spending Impacts

The estimated passenger spending activity at *Heartland Flyer* stations during FY 2009 is shown in Table 4. From these figures, it is estimated that *Heartland Flyer* passengers spent more than \$18 million on lodging, meals, shopping, and entertainment on their trips. Most of the passenger spending activity occurred at the route’s endpoints of Fort Worth (\$8.8 million) and Oklahoma City (\$5.9 million). This is not surprising as these are the two largest cities on the route and have major activity centers near the rail stations. Average passenger spending in communities along the route ranged from around \$50 per passenger in Pauls Valley to \$170 per passenger in Fort Worth. Oklahoma City and Norman were both computed to be around \$120 per passenger, while Gainesville was slightly over \$100 and Ardmore slightly under \$90 per passenger. No survey records were identified with Purcell as the destination station; as a result, the data report no passenger spending for that station.

Table 4 also reports the estimated sales tax revenues from passenger spending for each community served by the *Heartland Flyer*. Sales tax rates for each community were provided by the Oklahoma Tax Commission and the Texas State Comptroller of Public Accounts for stations in the respective states. Sales tax rates ranged from 8.25% in Fort Worth, Gainesville, and Norman to 9.0% in Pauls Valley. For

**TABLE 4 Impact of Spending and Sales Tax Revenue on Case Study Route, FY 2009**

Spending Station	Passenger Spending (\$)	Sales Tax Revenue (\$)
Oklahoma City	5,876,150	454,097
Norman	1,493,464	113,821
Pauls Valley	268,193	22,161
Ardmore	797,612	64,176
Gainesville	810,560	61,775
Fort Worth	8,786,451	669,637
Total	18,032,628	1,385,666

NOTE: No surveys reported Purcell as a destination for spending.  
SOURCE: 2009 *Heartland Flyer* passenger survey (11).

a given total spending level  $S$  and a local sales tax rate  $t$ , the total sales tax revenue  $R$  generated from spending by *Heartland Flyer* passengers in each community is estimated with the following equation:

$$R = t * \frac{S}{1 + t}$$

Using this equation to calculate the spending impacts assumes that the passenger's stated spending levels were reported on an after-tax basis, that is, that the amount reported was the total amount paid and not the (pretax) total cost of the items purchased during the trip. Researchers believed this was a good assumption given that the surveyed passengers would likely be able to recall the total spending including the tax more readily than the pretax price of the items purchased. The sales tax impact of the *Heartland Flyer* on the communities it serves appears to be substantial: it is estimated that purchases made by passengers resulted in total sales tax revenues of almost \$1.4 million. In general, the distribution of sales tax revenue was proportional to total spending, with Fort Worth and Oklahoma City receiving a majority (more than 80%) of the tax impacts. The state distribution of this sales tax revenue was \$731,412 in Texas (53%) and \$654,254 in Oklahoma (47%). This distribution raises an interesting discrepancy between the users of the service and the distribution of the resulting sales tax revenue. Specifically, while Texas residents accounted for approximately 20% of *Heartland Flyer* passengers, more than half of the sales tax revenue attributed to all passengers was generated in Fort Worth or Gainesville. However, in terms of potential return on state investment in the rail service, the costs of the route are shared equally between the two states. The estimated split of sales tax revenue between the two states is also approximately equal, implying an equal tax impact in each state, at least in the context of sales tax revenues from passenger spending.

### Potential Impacts from Loss of Rail Service

Perhaps a more critical policy question in the context of the passenger spending impacts of intercity passenger rail service is the estimated loss of spending and sales tax revenue if the rail service were to be discontinued. One question on the survey asked the passenger to report how he or she might make the trip if the service did not exist, with automobile, air carrier, or intercity bus as options. An option for "would not make the trip" was provided for passengers who might forgo their trips in the absence of the service. If the *Heartland Flyer* rail service were discontinued, approximately 70% of passengers would divert their trip to another mode, and the balance would not make a trip at all. While the 70% of passengers who diverted from other modes have measurable benefits for regional mobility (e.g., an estimated reduction of 7.9 million annual vehicle miles traveled on parallel roadways), the focus of this analysis is the spending attributed to the 30% of passengers who would forgo their trip, or induced travelers. In the context of passenger spending, any spending by induced travelers (more than 20,000 passengers in FY 2009) would also be lost if the service were discontinued.

Analysis of survey responses found that induced travelers reported higher median spending levels (\$153.51 during the average season, \$187.42 during the peak season) than travelers who would divert to other modes (\$112.30 and \$161.76, respectively). While the differences in spending levels between diverted and induced passengers were not significant (average  $\chi^2 = 9.38$ ,  $p = .0523$ ; peak  $\chi^2 = 2.998$ ,  $p = .5582$ ), the difference in medians suggests that discontinuing the *Heartland Flyer* would, on a per visitor basis, have a larger impact

TABLE 5 Estimated Spending and Sales Tax Revenue Losses If Service Were Discontinued

Spending Station	Passenger Spending (\$)	Sales Tax Revenue (\$)	Percentage of Total
Oklahoma City	1,979,506	152,972	34
Norman	258,240	19,681	17
Pauls Valley	63,031	5,204	24
Ardmore	39,766	3,200	5
Gainesville	391,615	29,846	48
Fort Worth	2,309,163	175,987	26
Total	5,041,321	386,890	28 (average)

NOTE: No surveys reported Purcell as a destination for spending.  
SOURCE: 2009 *Heartland Flyer* passenger survey (11).

on the communities it serves, since the lost revenue from induced trips is higher than the trips that would shift to other modes in the absence of the service.

Table 5 reports the estimated spending and sales tax revenue that would be lost if the *Heartland Flyer* service were to be discontinued. These figures were developed through the same basic procedure as the data reported in Table 4. Also reported in Table 5 are the estimated potential lost passenger spending and sales tax revenue as a percentage of the total estimated spending and tax revenue for each community. It is estimated that more than \$5 million of direct spending by current *Heartland Flyer* passengers would not transfer to other modes if the service were discontinued, with an estimated loss of \$386,890 in local sales tax revenue. On average, each induced passenger contributes an estimated \$11.31 of local sales tax revenue in communities served by the train. Discontinuing the *Heartland Flyer* service would have a significant impact on the local economies of the communities served by the route, with nearly 30% of the total spending lost under this scenario. Almost half of the revenue attributed to the passenger rail service in Gainesville, for example, would be lost. The distribution of induced traveler spending and sales tax revenue is approximately equal to the distribution of total revenue reported in the previous section, with approximately 53% to Texas communities and 47% to Oklahoma communities.

The economic impacts of passenger rail service reported in this section come with several important caveats. It is likely that some travelers estimated spending levels, as the actual spending had not occurred at the time the survey was administered. It is suspected that for these passengers, the actual spending levels were greater than what was estimated and reported on the survey. Not all estimated spending took place immediately in the communities served by the *Heartland Flyer*. It is impossible to know exactly where the spending occurred; for example, the data do not indicate how much was spent in the local area (i.e., the area surrounding the station) and how much was spent elsewhere in the region surrounding the communities with stations. Instead, the figures reported are likely good approximations of the true average spending levels. The sales tax revenue computations only consider sales tax and do not include hotel occupancy or other special taxes that a community might impose. Including the revenue projections from these taxes would likely increase the overall tax generation attributed to the *Heartland Flyer* service. Finally, the computations only address the direct economic impacts related to passenger spending and do not consider any multiplier effects from the spending. If these effects were incorporated, they would likely increase the impacts of the *Heartland Flyer* service beyond what was identified in this analysis.

## DISCUSSION AND CONCLUSIONS

The case study described in this paper reports the findings of an onboard survey of passengers using the *Heartland Flyer*, an intercity passenger rail route that serves communities in Oklahoma and Texas. Specific findings from the survey on passenger spending during the rail trip on lodging, meals, shopping, and entertainment were used to measure the economic impacts of the rail service. This direct measurement approach to identify passenger expenditures and the resulting community impacts is somewhat rare in this context, having only been applied in one other study of existing intercity rail passengers. The analysis examined total passenger spending as well as patterns in median spending levels among different passenger trip purposes and trip durations. Total community spending impacts attributed to *Heartland Flyer* passengers were estimated at more than \$18 million, which generated an estimated \$1.39 million in sales tax revenue to communities served by the train. Discontinuing the train service would result in an estimated 30% reduction in these impacts.

The findings of this study can be applied in a number of intercity passenger rail policy and planning contexts. For policy development, these findings clearly demonstrate the value of intercity passenger rail in the communities and states in which they operate, both in the study corridor as well as similar corridors across the country. It is evident from this study that the impacts of intercity passenger rail service extend beyond the basic provision of mobility alternatives in an intercity corridor and that a substantial amount of economic activity is generated by rail passengers visiting communities along the route. Metrics such as the median passenger spending levels and the total sales tax revenue generated by rail passengers are useful when discussing the impacts of intercity passenger rail with legislators and transportation agency executives who are tasked with developing transportation policy and investment strategies for their jurisdictions while also faced with severe budget crises. From the planning and service development perspective, two key applications emerge. The findings of this analysis can be used by state rail planners to strengthen applications for federal matching funds for intercity passenger rail capital investment programs created by PRIIA. Criteria for project selection under the grant program created in PRIIA Section 301, for example, direct that greater consideration be given to projects with "anticipated positive economic and employment" impacts (1). Similar language is included in PRIIA Section 303 regarding the inclusion of passenger rail projects in a state rail plan. The community-level spending impacts could also be used by rail planners to support efforts to obtain stakeholder buy-in for new passenger rail projects in local communities. In a policy climate with a renewed emphasis on transparency and accountability in the use of public resources, directly measured impacts may be particularly helpful.

Given the wealth of information obtained from the onboard survey direct measurement approach used in this analysis, future research studies should seek to improve the technique. Some effort should be made, for example, to develop a standard group of onboard survey questions to identify rail passenger spending. Such a development could provide the basis for a methodology in which corridor-by-corridor results could be directly compared. The reader is cautioned that the findings of the present study should not be generalized to other existing or proposed intercity passenger rail lines without first assessing the similarities between the *Heartland Flyer* corridor and the corridor in question. A similar approach could be taken to identify the economic impacts of other intercity travel modes such as intercity bus, which would allow for a comparison between the economic impacts of these modes and guide policy development accord-

ingly. With a standard methodology, onboard passenger surveys of other intercity rail corridors across the country could be undertaken to provide a clear assessment of the economic impacts of the nation's intercity passenger rail network as it exists today and to provide baseline data for future studies examining the economic impacts of new system investments.

## ACKNOWLEDGMENTS

The authors thank the Southwest University Transportation Center for providing financial support for the research study described in this paper and the anonymous reviewers for their helpful comments on the draft paper.

## REFERENCES

1. Federal Railroad Administration. *Overview, Highlights, and Summary of the Passenger Rail Investment and Improvement Act of 2008 (PRIIA)*. March 2009. [www.fra.dot.gov/Pages/2172.shtml](http://www.fra.dot.gov/Pages/2172.shtml). Accessed July 29, 2010.
2. Federal Railroad Administration. *A Vision for High Speed Rail in America*. April 2009. [www.fra.dot.gov/Downloads/RRdev/hrrspfacts.pdf](http://www.fra.dot.gov/Downloads/RRdev/hrrspfacts.pdf). Accessed July 29, 2010.
3. Weisbrod, G., and B. Weisbrod. In *Transportation Research Circular 477: Assessing the Economic Impact of Transportation Projects: How to Choose the Appropriate Technique for Your Project*. TRB, National Research Council, Washington, D.C., 1997.
4. *Midwest Regional Rail System: A Transportation Network for the 21st Century*. Executive summary report. Transportation Economics & Management Systems, Inc., and HNTB Corporation, Frederick, Md., 2004.
5. TranSystems, and ESH Consult. *Chicago to St. Louis 220 mph High-Speed Rail Alternative Corridor Study, Volume 2: Ridership and Benefits*. Midwest High-Speed Rail Association, Chicago, Ill., 2010.
6. Lynch, T. *Florida High-Speed Ground Transportation Economic Benefit and Cost Impact Restudy*. Econometrics Consultants, Inc., and Center for Economic Analysis, Florida State University, Tallahassee, 2002.
7. Brand, D., M. R. Kiefer, T. E. Parody, and S. R. Mehndiratta. Application of Benefit-Cost Analysis to the Proposed California High-Speed Rail System. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1742, TRB, National Research Council, Washington, D.C., 2001, pp. 9–16.
8. *Report to the Legislature: December 2009*. California High-Speed Rail Authority, Sacramento, Calif., 2009.
9. Economic Development Research Group, Inc., and KKO Associates. *Final Report: Economic Benefits of Amtrak Downeaster Service*. Maine Department of Transportation, Augusta, 2005.
10. Taylor, J. C., H. Singh, and P. Isely. *Michigan Passenger Rail Station Community Benefits Study*. Report. Grand Valley State University, Grand Rapids, Mich., 2009.
11. Sperry, B., and C. Morgan. *Measuring the Benefits of Intercity Passenger Rail: A Study of the Heartland Flyer Corridor*. Publication SWUTC/10/169116-1. Southwest University Transportation Center, Texas Transportation Institute, Texas A&M University, College Station, 2010.
12. *Amtrak System Timetable Fall 2009–Winter 2010*. National Railroad Passenger Corporation, Washington, D.C., 2009.
13. Carter-Burgess, Inc. *The Heartland Flyer: Oklahoma's Passenger Rail Service: Economic Benefit Report*. C&B Project No. 020874.010. Rail Programs Division, Oklahoma Department of Transportation, Oklahoma City, 2005.
14. Sperry, B., and C. Morgan. Evaluating the Impacts of Passenger Rail Service: Case Study and Lessons Learned. *Proc., 2010 Joint Rail Conference*, Urbana, Ill., 2010.

*The views expressed in this paper are the opinions of the authors and do not represent the official views of the Texas Transportation Institute, the Oklahoma Department of Transportation, the Texas Department of Transportation, or Amtrak. Any errors, inaccuracies, or omissions are the responsibility of the authors.*

*The Intercity Passenger Rail Committee peer-reviewed this paper.*

# Cluster Analysis of Intercity Rail Passengers in Emerging High-Speed Rail Corridor

Benjamin R. Sperry, Kristopher D. Ball, and Curtis A. Morgan

Recent pledges by the U.S. federal government to increase investments in high-speed passenger rail have highlighted the importance of identifying the potential impacts of these investments on current rail passengers. Identification of these impacts depends largely on the demographics, travel habits, and attitudes of these passengers—in essence, who these passengers are. This investigation applied cluster analysis, a method typically used to determine potential consumer groups for marketing research, to identify similar groups of rail passengers riding Amtrak's *Hiawatha* service between Chicago, Illinois, and Milwaukee, Wisconsin. These groups were then analyzed for their potential responses to future high-speed rail service. Analysis results showed that clustering was an effective method to determine differences between segments of a market, in this case, current passenger rail users. Further analysis indicated that these segments exhibited potentially different responses to future high-speed rail service. It is recommended that attention be paid to the potential needs of these current passenger segments to increase the likelihood of success for future high-speed rail service.

Recent policy and regulatory developments have stimulated interest in high-speed passenger rail among planners, policy makers, and the general public in the United States. Policies created in the Passenger Rail Improvement and Investment Act of 2008 established new roles for the federal government in the development of high-speed intercity passenger rail service (1). Funding of grant programs established in this act and appropriations received through the *American Recovery and Reinvestment Act of 2009* include a substantial federal match for state investments in high-speed intercity passenger rail. These events signal a new direction in U.S. transportation policy by providing for the development of passenger rail corridors on a playing field that is arguably more level with other modal funding programs than at any other point in the history of U.S. transportation policy (2).

As the United States moves forward with the redevelopment of its long-neglected intercity passenger rail network, a number of questions and challenges remain to be faced by rail planners and pol-

icy makers, who are tasked with developing capital investment programs and service development plans for new passenger service. In emerging U.S. high-speed rail corridors where intercity passenger rail service is currently offered by Amtrak or another operator, the development of new high-speed lines is likely to have a profound impact on passengers using the existing service. Arguably, current rail passengers along a proposed high-speed rail route represent a great ridership potential for the new service, as these passengers are at least familiar, if not accustomed, to the concept of passenger rail for intercity trips. Amtrak corridor or long-distance trains operate over some portion of all 10 federally designated high-speed rail corridors. As a result, the impact of new high-speed rail service on current rail passengers in a corridor is not a trivial question. This study examines the impacts of new high-speed rail development from the perspective of travelers who already use existing conventional intercity rail services in an emerging U.S. high-speed corridor between Chicago, Illinois, and Milwaukee, Wisconsin. Five passenger groups are identified by using cluster analysis, and from the characterization of each cluster, potential response to proposed high-speed rail service is projected.

## LITERATURE REVIEW

Cluster analysis is broadly defined as a family of multivariate analysis tools that can be used to form groups of individuals or objects from a larger population. The primary purpose of such analyses is to use these tools to form groups (known as clusters) from a set of multivariate data in which a group structure may otherwise not be readily apparent to the analyst. A number of techniques and algorithms can be used in a cluster analysis, but the main computational goal is to identify clusters from the data set such that the similarity among objects within a cluster is maximized while the similarity among groups is minimized (3, 4). Cluster analysis is used extensively in marketing research and product evaluation activities as an alternative to the use of traditional grouping means such as demographic categories (5).

Given its extensive use in marketing research and evaluation, it is not surprising that cluster analysis has been adopted by transportation planners as a means of better understanding transportation system users and gauging the impacts of system changes on these users. Clustering has been identified as one of several analysis tools available for the market segmentation of transit passengers, with the primary objective in such analyses being to develop detailed marketing and promotion strategies for various rider groups (6). For example, clustering has been used to identify groups of transit passengers in

---

B. R. Sperry and C. A. Morgan, Texas Transportation Institute, Texas A&M University System, 3135 TAMU, College Station, TX 77845-3135. K. D. Ball, URS Corporation, 1 Penn Plaza, Suite 610, New York, NY 10119-0698. Corresponding author: B. R. Sperry, b-sperry@ttimail.tamu.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 31–38.  
DOI: 10.3141/2261-04

San Mateo County, California (7), and Salt Lake City, Utah (8), and ferry passengers in San Francisco, California (9). Cluster analysis has also been used to identify the potential for switching from personal vehicle to transit or other alternatives based on attitude theory (10), gender differences (11), and residential location and lifestyle behaviors (12).

Evidence of cluster analysis for intercity passenger rail planning is limited in the literature. A 1992 study by Pas and Huber used clustering in an analysis of data obtained from surveys of potential intercity passenger rail travelers in the Piedmont region of North Carolina (13). Using the FASTCLUS nonhierarchical, agglomerative analysis algorithm, they identified five segments of potential rail travelers: the functional traveler, the day tripper, the train lover, the leisure-hedonic traveler, and the family traveler. They concluded that the segmentation of potential rail travelers via clustering provided a greater insight into the preferences and needs of these travelers and that the clustering approach was more useful than a priori clustering (i.e., business and nonbusiness) in identifying potential market segments.

More recently, Dallen, in the context of rail planning for sustainable tourism, applied clustering to surveys of travelers on two branch line railways in the Cornwall region of Southwest England. In one study, Dallen analyzed attitudes of travelers on the Looe Valley Branch Line Railway by using complete linkage agglomerative hierarchical clustering. He identified five user segments: train devotees (representing 22% of journeys on the line), infrequent enthusiasts (15%), train tolerators (29%), contented car users (24%), and last resort riders (10%) (14). In his second study, Dallen clustered both rail users and nonusers on the St. Ives Branch Line by using the same analysis technique and identified three clusters of rail users: road regulars (45%), public transport reliant (11%), and train enjoyers (44%) (15). In both studies of rail passengers in a developing tourist region, Dallen demonstrated that cluster analysis helped to uncover information about how passengers use the rail services that otherwise would have been masked by traditional traveler definitions. These studies demonstrate that clustering can be an effective way to reveal information about the traveling public that may not otherwise be readily apparent to the investigator.

## RESEARCH PROBLEM

While the application of clustering has proved useful in a number of modal contexts, the authors are not aware of any analyses that have used clustering to understand current rail passengers on intercity corridors in the United States. This would include the application of clustering to understand the effects of new high-speed rail development on current rail passengers. Although ridership studies for proposed high-speed rail systems in California (16) and the Midwest (17) have obtained passenger survey data from current rail travelers in affected corridors, these efforts were primarily for establishing modal diversion rates or other parameters for detailed ridership forecasts. Furthermore, these studies used traditional or a priori passenger groupings (i.e., business and nonbusiness travelers) that may only mask actual differences between groups. Consequently, this paper aims to contribute to passenger rail planning by performing a cluster analysis of survey data obtained from current rail passengers in an emerging high-speed rail corridor and, using the findings of the analysis, estimate the response of each group to proposed high-speed rail service.

## DATA AND METHODS

### Case Study Corridor

The emerging high-speed passenger rail corridor examined in this paper runs between Chicago and Milwaukee. Currently operating along the 86-mi (138-km) corridor is the *Hiawatha* service, a passenger rail route operated by Amtrak in partnership with the Illinois Department of Transportation and the Wisconsin Department of Transportation. Intermediate stops include Glenview in Illinois and Sturtevant and Milwaukee's General Mitchell International Airport in Wisconsin. Amtrak's current timetable reports the endpoint-to-endpoint travel time to be 89 min. Service frequencies Monday through Saturday are seven trains in each direction, with six on Sundays (18). Ridership during the most recent FY (October 2009 to September 2010) totaled 783,057 passengers and has grown 50% over the past 6 years. The Chicago-Milwaukee corridor is a segment of the Chicago-Minneapolis-Saint Paul spoke on the emerging high-speed passenger rail Chicago hub network. This corridor, along with other corridors in the Midwest, is being developed as part of a proposed multistate rail system known as the Midwest Regional Rail System. When this regional rail system is fully implemented, service plans call for 17 high-speed daily passenger trains in each direction on the Chicago-Milwaukee corridor, with projected top speeds of 110 mph (177 km/h) on the route (17).

### Data Source

The data analyzed in this study were obtained from an onboard survey of *Hiawatha* service passengers conducted on all corridor trains during an average weekday (Thursday) and weekend (Saturday) in March and October of 2005. The four-page survey, which used a hand-out-hand-back distribution and collection scheme, included 24 questions related to the passenger's trip, evaluation of onboard service, preferences for service configurations, and demographics. All *Hiawatha* service passengers during the survey dates were eligible to participate and were offered the opportunity to complete a survey form. A total of 3,191 surveys were obtained from 5,204 passengers boarding the *Hiawatha* service during the 4-day data collection period, resulting in a participation rate of 62%. No formal tabulations were maintained on the nature of refusals. Since data collection only took place for a total of 4 days in March and October, some seasonal variation in passenger characteristics likely exists that was not captured with these data. While a full 12 months of data would have been ideal, the cost of such an undertaking was well beyond the budget. A more detailed examination of the onboard survey data collection and data set preparation and a preliminary analysis of these data are included in a 2010 conference paper (19).

### Method of Analysis

Cluster analysis was used to identify groups of intercity rail passengers who use existing, conventional intercity passenger rail service from onboard survey data of passengers in an emerging high-speed rail corridor. A formal cluster analysis requires the analyst to make three important decisions: first, the clustering analysis approach and specific algorithm must be selected; second, the variables used must be identified; finally, the number of clusters must be specified. To develop the rail passenger cluster groups, this analysis employed a

hierarchical clustering technique with the Ward's method hierarchical clustering algorithm, one of the most popular methods for clustering analysis (3). The Ward's method was selected for this analysis because it minimizes the distance between units within an individual cluster (using a sum-of-squares criterion).

Eleven clustering variables were used, three representing the characteristics of the passenger's overall travel on the rail service, three representing the passenger's likelihood of increased rail travel frequency in response to service changes, and five demographic characteristics. The three variables used to characterize the passenger's rail travel were day type (weekday or weekend), round trip that day (yes or no), and passenger frequency of rail travel in the corridor (five options with a range of monthly usage). The three variables representing the passenger's likelihood of increasing his or her frequency of rail travel examined the likelihood of increased frequency resulting from faster trip time, more daily departures, and lower fares measured on a three-point scale (very likely, somewhat likely, and not likely). The five demographics variables were passenger gender, age, highest level of education attained, number of household vehicles, and annual household income. Only records with valid responses for all 11 clustering variables were considered in the analysis. The remaining 13 questions from the survey instrument were not used in the cluster analysis because they were not suitable for clustering (i.e., contained nominal scales) or did not relate to the purpose of the analysis (e.g., use of ticket discount programs). From the original sample of 3,191 passenger surveys, 2,225 surveys met this requirement, or approximately 79% of the records. As a result, the effective sampling percentage for the purposes of the clustering analysis was 42.8% of all passengers who rode *Hiawatha* service trains during the survey period.

The final analysis consideration was the selection of the number of clusters that would be obtained from the analysis. While there is no formal method or algorithm for selecting the number of clusters, the general principle is that the number of clusters should not be so small that important differences between cluster members are blurred, but also not so large that the results are not practical for application. One author noted that the selection of the number of clusters should be guided by a combination of four factors: statistical output, the transferability of the output, lessons from past research efforts, and researcher intuition and common sense (12). The hierarchical clustering analysis routine in the SPSS software program was used to identify cluster membership for each of the 2,225 records for cluster solutions for four to 10 clusters. A core group of three clusters was identified that remained stable for cluster solutions from four to seven. Two other stable clusters were identified, resulting in a final solution that consisted of five clusters of passengers.

## RESULTS OF CLUSTER ANALYSIS

The first objective of this study was to use clustering analysis methodology to examine the *Hiawatha* service onboard passenger survey data. The analysis resulted in a solution that consisted of five clusters. The next phase of the analysis was to examine the characteristics of these five clusters and to identify the critical variables that make each passenger group unique.

### Cluster Characterization

Identification of individual clusters within the final five-cluster solution was undertaken with a detailed analysis of the following dimensions:

- Demographics,
- Travel behavior,
- Reasons for choosing the train, and
- Likelihood of increased rail patronage in response to service changes.

Using these findings, trends and patterns of the members of each of the five clusters were identified. Based on these trends and patterns, the five distinct passenger groups were identified and inspired the following names: convenient casuals, operative regulars, functional affluents, fortuitous flyers, and mobile youth. Table 1 provides a summary description of each cluster and the relative size of each cluster when percentages are applied to the total annual ridership (783,057 passengers) from FY 2010.

Convenient casuals are the largest cluster, representing 34% of passengers. Convenient casuals are characterized by a lower frequency of rail trips for a wide array of functions. Conversely, operative regulars (16%) are characterized by a higher than average percentage of trips for one specific purpose, trips to and from employment. Functional affluents are characterized by relatively high incomes and use of passenger rail services for a variety of trip purposes. These passengers comprise the second-largest group, representing 25% of passengers. Fortuitous flyers (14%) are largely low-income travelers who are relatively limited in their transportation options compared with the other clusters. Finally, the smallest cluster, the mobile youth, primarily use the rail service because of its convenience. This cluster, which represents 11% of passengers, has relatively more transportation options than those in other clusters and is more likely to travel in groups.

The demographic and travel behavior profiles of each cluster based on the identifying dimensions (demographics, travel behavior, reasons

TABLE 1 Passenger Clusters from 2005 *Hiawatha* Service Passenger Survey Data

Cluster (% of passengers)	Size (FY 2010)	Description
Convenient casuals (34)	266,239	Casual travelers who use passenger rail services infrequently for wide array of functions and purposes
Operative regulars (16)	125,289	Passenger rail travelers who use rail services frequently, typically for employment
Functional affluents (25)	195,764	High income travelers who use passenger rail services for specific functions
Fortuitous flyers (14)	109,628	Passenger rail users with lower incomes and limited transportation options who use services for an array of functions
Mobile youth (11)	86,137	Passenger rail users with a number of transportation options who travel in groups largely to and from educational establishments

for choosing the train, and likelihood of increased patronage) are discussed in the next two subsections.

### Demographic Profiles

Table 2 provides the demographic profile of each cluster. The convenient casuals, fortuitous flyers, and mobile youth were typically female. The operative regulars were largely male (only 35% female), compared with 51% female for all passengers. The median age for all passengers was 38 years. Convenient casuals, operative regulars, and functional affluents all reported a median age within 10 years of the samplewide median. Functional affluents reported a higher median age (45 years) than the total group of passengers, and fortuitous flyers and mobile youth reported a median age (23 years) that was substantially lower than the other clusters.

A majority of convenient casuals, operative regulars, and functional affluents had earned at least a bachelor's degree, while some college was the highest level of education attained by most of the fortuitous flyers and mobile youth. This finding, along with the median age of the latter two groups, supports earlier assumptions that the fortuitous flyers and mobile youth groups were largely composed of college students or young adults. Graduate degrees had been obtained by 39% of the operative regulars and 50% of the functional affluents compared with 28% of all passengers. This finding is consistent with the higher median age of these clusters as compared with all passengers.

Except for mobile youth, the reported annual household income and highest educational attainment of each cluster generally followed

the same trend. With respect to income, the mobile youth discontinued their similarity to the fortuitous flyers. Mobile youth had a median household income of \$90,000 compared with a median household income of \$11,000 for fortuitous flyers, suggesting that fortuitous flyers may face additional limitations in their transportation options. Since a large percentage of mobile youth were seemingly still in college, one might wonder if these passengers were reporting their overall family incomes rather than their personal incomes. Both operative regulars and functional affluents reported annual median household incomes over \$120,000, which was also in line with the higher median educational levels of these clusters relative to all passengers. The demographics of all five clusters show that the sample population covered a wide range of age, income, and educational levels.

Table 3 contains information on the travel behavior characteristics of each of the five clusters. Operative regulars had a substantially larger percentage of weekday travelers (93%) than the pool of all passengers (53%). Operative regulars also were characterized by a high number of same-day round-trip travelers (94%) compared with the pool of all passengers (51%). The mobile youth exhibited a low percentage of day round-trip behavior (24%) compared with the pool of all passengers.

The operative regulars and fortuitous flyers stand out for having high percentages of travelers associated with one or two trip purposes. Within the operative regulars group, 83% of the surveyed passengers had a primary trip purpose of commuting to or from work. The high proportion of work commuters among the operative regulars is unsurprising given this group's high propensity for daily round-trip

TABLE 2 Demographic Profile of Passenger Clusters

Variable	Convenient Casuals ( <i>n</i> = 751, 34%)	Operative Regulars ( <i>n</i> = 350, 16%)	Functional Affluents ( <i>n</i> = 563, 25%)	Fortuitous Flyers ( <i>n</i> = 304, 14%)	Mobile Youth ( <i>n</i> = 257, 11%)	Total ( <i>n</i> = 2,225, 100%)
Female (%)	58	35	49	55	53	51
Age						
Median (years)	41	41	45	23	23	38
Younger than 18 (%)	0	0	0	8	11	2
18 to 24 (%)	11	3	3	60	60	20
25 to 34 (%)	27	24	15	27	13	22
35 to 44 (%)	19	37	30	4	16	23
45 to 54 (%)	20	28	31	0	1	19
55 to 64 (%)	16	7	17	0	0	11
65 and older (%)	7	0	4	0	0	3
Highest level of education (%)						
Some high school	1	1	0	9	11	3
High-school graduate	10	3	0	15	19	8
Some college	15	5	1	40	67	19
Associate's degree	8	6	6	4	3	6
Bachelor's degree	42	46	43	25	1	36
Graduate degree	24	39	50	7	0	28
Household income						
Median	\$57,100	\$124,400	\$145,700	\$11,000	\$90,000	\$80,400
Less than \$10,000 (%)	1	1	0	47	0	7
\$10,000 to \$19,999 (%)	3	1	0	28	0	5
\$20,000 to \$29,999 (%)	6	0	0	15	0	4
\$30,000 to \$39,999 (%)	13	3	0	9	1	6
\$40,000 to \$49,999 (%)	17	5	0	1	9	8
\$50,000 to \$74,999 (%)	36	11	3	0	26	17
\$75,000 to \$99,999 (%)	16	14	12	0	25	13
\$100,000 to \$149,999 (%)	7	33	38	0	19	19
\$150,000 to \$199,999 (%)	1	17	17	0	6	8
\$200,000 or more (%)	1	16	29	0	16	12

TABLE 3 Travel Behavior Profile of Passenger Clusters

Variable	Convenient Casuals (n = 751, 34%)	Operative Regulars (n = 350, 16%)	Functional Affluents (n = 563, 25%)	Fortuitous Flyers (n = 304, 14%)	Mobile Youth (n = 257, 11%)	Total (n = 2,225, 100%)
Weekday (%)	41	93	58	39	42	53
Round-trip that day (%)	47	94	51	33	24	51
Trip purpose (%)						
Business	11	5	25	1	4	11
Education	4	3	4	18	38	10
Leisure or entertainment	18	1	19	13	12	14
Personal business	5	5	4	3	2	4
Shopping	3	1	6	3	5	4
Visit family or friends	48	3	21	58	35	34
Work commute	11	83	20	4	3	24
Trip frequency (%)						
Less than one per month	67	0	72	71	69	58
One to four per month	27	5	22	25	26	22
Five to nine per month	6	10	6	3	3	6
10–19 per month	1	29	1	1	1	5
20 or more per month	0	55	0	0	1	9
Alternative travel mode (%)						
Airplane	5	1	5	7	9	5
Bus	11	4	4	23	13	10
Drive car or truck	53	48	66	33	41	52
Passenger in car or truck	12	4	12	19	22	13
Rental car	6	1	5	3	3	4
Would not make trip	11	31	7	15	11	14
Metra commuter rail	1	11	1	1	1	3
Traveling (%)						
Alone	66	97	60	66	60	69
With family	21	2	27	18	25	19
With a group	14	1	14	16	15	12
Household vehicles (%)						
Zero	14	2	4	34	6	12
One	35	16	13	31	7	23
Two	38	62	51	14	37	42
Three	8	16	21	11	30	15
Four or more	5	4	11	10	20	9

travel, high incomes, and high educational attainment. Among the fortuitous flyers, 76% of travelers were either visiting family and friends or traveling for educational purposes (58% and 18%, respectively). The mobile youth cluster reestablished its trend of similarity to the fortuitous flyers with 73% of its passengers traveling either to visit family and friends or for educational purposes (35% and 38%, respectively). Information for the fortuitous flyers and mobile youth support earlier inferences that they comprised a fairly large percentage of students. All clusters except for operative regulars took trips infrequently (less than one trip per month). In contrast, more than half the operative regulars used the *Hiawatha* service 20 or more times per month.

### Travel Behavior Profiles

Examining alternative travel mode behavior reveals how passengers might travel if rail service were discontinued, either on a temporary basis (i.e., because of weather or labor issues) or if the funding support for the service was unavailable. While the pool of all passengers had a very low percentage of travelers who would divert trips to air travel (5%), the operative regulars had an even lower percentage (1%). Further, this group was also the least likely to rent a vehicle as an

alternative for their trip. Conversely, 9% of mobile youth would divert to an airplane if passenger rail were unavailable. This suggests that these passengers may have longer travel distances or less time-sensitive trips than passengers in other clusters. Fortuitous flyers had a higher propensity to divert their trips to bus (23% versus 10% average), and they had a lower than average percentage of drivers who would divert trips to driving (33% versus 52% average). The fortuitous flyers and mobile youth were also more likely to make their trips as a passenger in a vehicle than the pool of all passengers (19% and 22%, respectively, compared with 13% average). These findings support earlier findings that fortuitous flyers have limited transportation options and that mobile youth find convenience in rail and are less likely than other clusters to drive a vehicle to make trips. Operative regulars remained consistent with their commuter characterization as they had the highest percentage of passengers who would use Metra commuter rail services as an alternative (11% compared with 3% for the pool of all passengers). This finding suggests that a large number of these passengers commuted from locations within Wisconsin to Chicago, the hub of the Metra network. Of operative regulars, 31% reported “would not make trip” as their primary alternative to rail service. One might not expect passenger groups such as the operative regulars (making a high percentage of work commutes) to forgo their trips. The reasons for this trend, however, were not

apparent from the data. Future surveys should seek to identify these reasons.

The operative regulars were more likely to be traveling alone than the pool of all passengers, as many of them make their daily commutes via the *Hiawatha* service (97% versus 69% average). The functional affluents and mobile youth were more likely than the pool of all passengers to be traveling with family (27% and 25%, respectively, versus 19%). A stark contrast was evident between fortuitous flyers and mobile youth in the number of household vehicles for each passenger, with over half of the fortuitous flyers having zero or one household vehicle and over half of the mobile youth reporting three or more vehicles. This trend, coupled with the alternative travel mode behavior of the fortuitous flyers, further supports the assumption that these passengers have limited transportation options and may be more reliant on the current rail service than other clusters. Very different trends in these categories by the mobile youth suggest that these passengers may be less dependent on the current rail service than fortuitous flyers, despite being less inclined to drive themselves compared with the other three clusters.

### Reasons for Choosing Rail

Table 4 reports the three most frequently cited reasons each cluster gave for choosing rail for the current trip. The most frequently listed reason among all clusters was "avoid traffic or parking problems," which was cited by nearly 70% of travelers. This is not surprising, as expressways within this corridor can become congested during the day and parking in Chicago can be costly, making rail an attractive travel option for corridor trips. Another frequently listed reason for using passenger rail services was that rail service was "faster than other options." This was listed in the top responses for four clusters. Among the operative regulars this response was particularly high (32%). These passengers could potentially have more time-

TABLE 4 Three Most-Frequent Reasons for Choosing Passenger Rail Service by Passenger Cluster

Cluster (% of passengers)	Top-Three Reasons Cited
Convenient casuals (34)	Avoid traffic or parking problems (70%) Convenient to destination (19%) Faster than other options (17%) Less expensive than other options (17%) More comfortable than other options (17%)
Operative regulars (16)	Avoid traffic or parking problems (81%) Opportunity to read, work, or sleep (34%) Faster than other options (32%)
Functional affluents (25)	Avoid traffic or parking problems (84%) Opportunity to read, work, or sleep (27%) Convenient to destination (20%)
Fortuitous flyers (14)	Avoid traffic or parking problems (39%) Less expensive than other options (26%) Faster than other options (20%) Convenient to destination (20%)
Mobile youth (11)	Avoid traffic or parking problems (49%) Convenient to destination (29%) Faster than other options (26%) Less expensive than other options (26%)
All trips (100)	Avoid traffic or parking problems (69%) Faster than other options (21%) Opportunity to read, work, or sleep (20%)

sensitive trips as a large percentage use passenger rail services to make daily commutes.

Operative regulars were the only cluster not to have "convenient to destination" within their top responses. This finding, coupled with the fact that 31% of these passengers would not make their trips if passenger rail service were unavailable, suggests that for operative regulars the service is more about functionality (being an essential element in the commute to work) than convenience (improving the quality of the commute to work). Conversely, only convenient casuals listed both "more comfortable than other options" and "convenient to destination" within their top three reasons for choosing rail, suggesting that for these passengers convenience is more of an issue than functionality. "Less expensive than other options" was listed within the top reasons for choosing rail by convenient casuals, fortuitous flyers, and mobile youth, suggesting that price is an important factor for these groups. These travelers may be more inclined to move their trips to other modes with an increase in the price of passenger rail relative to other modes.

### Likelihood of Increased Patronage

Potential factors to increase frequency are listed in Table 5. Operative regulars are very likely to increase ridership with faster trip times and increased train frequencies, and they are also the second most likely to increase ridership with decreased fares. These findings are in agreement with prior conclusions regarding the functional nature of trips for operative regulars, as all of these changes improve the functional efficiency and convenience of rail service for passengers. Despite indicating a positive response to faster trip times, functional affluents were least likely to increase ridership with faster trips.

TABLE 5 Potential Service Changes to Increase Ridership by Passenger Cluster

Cluster (% of passengers)	Faster Trip Time (%)	Lower Fares (%)	Increased Frequency (%)
Convenient casuals (34)			
Very likely	55.9	72.7	57.8
Somewhat likely	28.9	21.3	33.0
Not likely	15.2	6.0	9.2
Operative regulars (16)			
Very likely	78.6	74.3	84.9
Somewhat likely	17.1	19.1	12.3
Not likely	4.3	9.2	2.9
Functional affluents (25)			
Very likely	53.3	48.5	60.4
Somewhat likely	30.7	34.3	27.9
Not likely	16.0	17.2	11.7
Fortuitous flyers (14)			
Very likely	63.5	83.2	61.2
Somewhat likely	25.3	12.8	31.6
Not likely	11.2	3.9	7.2
Mobile youth (11)			
Very likely	63.8	74.3	64.6
Somewhat likely	26.5	18.3	28.4
Not likely	9.7	7.4	7.0
All trips (100)			
Very likely	60.8	68.4	64.0
Somewhat likely	26.7	22.7	27.7
Not likely	12.5	8.8	8.3

This finding supports earlier conclusions that convenience of rail service is important to functional affluents. Table 5 also reveals that fortuitous flyers and functional affluents are most and least likely, respectively, to increase ridership with decreases in fares, supporting earlier conclusions regarding household income and reasons for choosing rail for both groups.

## POTENTIAL RESPONSE TO NEW HIGH-SPEED RAIL PASSENGER SERVICE

The second objective of this study was to use the information gleaned from the clustering to estimate the potential response to new high-speed passenger rail service for each of the passenger groups identified. As an emerging high-speed rail corridor, the Chicago–Milwaukee route is expected to have improved passenger rail service in the coming decade. Plans call for 17 daily trains in each direction with 110 mph (177 km/h) service along the corridor (17). With the profiles developed in the previous section for demographics, travel behaviors, motivations for choosing rail, and the likelihood of increased rail patronage in response to service changes, potential response to new high-speed rail service can be projected for each cluster as follows:

- **Convenient casuals.** These passengers are satisfied with the current rail service in the corridor and are generally unlikely to change travel behavior solely on the basis of the faster trip times or new frequencies (or both) that will be offered with new high-speed passenger rail service. Because convenient casuals make travel decisions based on the best available option for the trip they are planning, increases in speed or frequency of service will reinforce high-speed passenger rail as the most attractive option for these travelers. Consequently, they will patronize new service provided it meets their travel needs. Raising awareness of the new service among them will be critical.
- **Operative regulars.** These passengers are frequent users of corridor rail service, primarily for travel to and from work either daily or 2 to 3 days per week. They are also very supportive of the faster travel times that will be offered by new high-speed rail service. However, given that their current trips appear to have a high utility, they will use train service as long as it maintains its current configuration and service levels do not degrade. New high-speed service should benefit these passengers' daily routines; with expanding service, this group may experience growth in terms of its relative size.
- **Functional affluents.** Like convenient casuals, functional affluents are satisfied with current service levels, including the convenience to their final destination. One-quarter of functional affluents are traveling for business. Although they have the highest diversion rate (93% of their trips would be made on other modes if service were discontinued), they appear less inclined to increase their rail trip frequency with any changes to current service. Thus, continued awareness of the current *Hiawatha* service and future high-speed rail service will be important to maintaining their patronage. The business traveler segment (and potentially other segments) of this group would likely appreciate continued reliability of the service.
- **Fortuitous flyers.** For fortuitous flyers, the current train service represents a relatively affordable opportunity to travel across the greater Chicago–Milwaukee region with higher speed and greater comfort than other options. They are generally very low income and are likely to be sensitive to the fare increases that may accompany new high-speed rail service. Patronage of new high-speed service by this group is likely to be related to the fare structure and levels of

the new service. Increased fares would likely diminish the relative number of fortuitous flyers as they would seek alternatives that are better aligned with their income levels.

- **Mobile youth.** The mobile youth, despite being the smallest group in terms of size, represent future passengers of high-speed service in the corridor. They are young and have many options for mobility, but choose the rail service for trips across the region because they recognize its benefits. Mobile youth have a strong likelihood of patronizing the planned high-speed rail service. Since they are familiar and comfortable with the passenger rail product and concept, mobile youth are likely to be strong supporters of new high-speed rail service as they age.

## CONCLUSIONS

In emerging high-speed rail corridors with current intercity rail service, understanding the current rail passenger profile is important for understanding how the current ridership will respond to new service when it is implemented. The cluster analysis performed in this study identified five distinct passenger groups, revealing a wealth of information with a wide range of applicability for intercity passenger rail planning, both for current service and for future high-speed services. The most relevant finding for planning and marketing the current rail service is that each of the five groups is very different, not only in who they are but also how they use the current service. The challenge for planning is to provide a product that meets the diverse needs of these groups.

For high-speed passenger rail planning and policy, each identified cluster has distinct considerations. For example, marketing efforts by the operator of the rail service may encourage the convenient casuals and functional affluents to patronize the new high-speed rail service. The operative regulars and mobile youth are likely to be strong supporters of new service and, based on the present findings, could be poised to grow in relative size as new service is implemented. From a policy perspective, as investment decisions are rendered for new services, provisions must be made so that the mobility needs of the fortuitous flyers continue to be met. From a methods standpoint, this study demonstrated the use of cluster analysis in an intercity passenger rail application. Rail planners at public agencies or service operators are encouraged to consider the use of clustering in their ridership profile studies and marketing efforts. However, each cluster analysis is unique, and the clusters developed here are not likely to transfer to other rail lines; consequently, further studies should be undertaken on other intercity passenger rail routes across the country to attempt to establish passenger clusters for other routes. To benefit service design, future studies should also seek to form clusters of nonpassengers along the route to better understand their travel behavior and how to make rail travel more appealing to these groups. Clustering may not replace traditional ridership analysis, but it provides another tool to analyze data in a different manner to complement other methods and provide more information for decision makers.

## ACKNOWLEDGMENTS

The authors thank Ron Adams, Ethan Johnson, and Arun Rao of the Wisconsin Department of Transportation for providing access to the data analyzed in this paper. The authors are also grateful for the helpful comments provided by the anonymous reviewers.

## REFERENCES

1. Federal Railroad Administration. *Overview, Highlights, and Summary of the Passenger Rail Investment and Improvement Act of 2008 (PRIIA)*. 2009. [www.fra.dot.gov/Pages/2172.shtml](http://www.fra.dot.gov/Pages/2172.shtml). Accessed July 29, 2010.
2. Federal Railroad Administration. *A Vision for High-Speed Rail in America*. 2009. [www.fra.dot.gov/Downloads/RRdev/hsrspfacts.pdf](http://www.fra.dot.gov/Downloads/RRdev/hsrspfacts.pdf). Accessed July 29, 2010.
3. Hair, J. F., W. C. Black, B. J. Babin, and R. E. Anderson. *Multivariate Data Analysis*, 7th ed. Prentice Hall, Upper Saddle River, N.J., 2010.
4. Everitt, B. S., S. Landau, and M. Leese. *Cluster Analysis*, 4th ed. Oxford University Press, New York, 2001.
5. Punj, G., and D. W. Stewart. Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research*, Vol. 20, 1983, pp. 134–148.
6. Elmore-Yalch, R. *TCRP Report 36: A Handbook: Using Market Segmentation to Increase Transit Ridership*. TRB, National Research Council, Washington, D.C., 1998.
7. Zhou, Y., K. Viswanathan, Y. Popuri, and K. E. Proussaloglou. Transit District Customers in San Mateo County, California: Who, Why, Where, and How. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1887, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 183–192.
8. Shiftan, Y., M. L. Outwater, and Y. Zhou. Transit Market Research Using Structural Equation Modeling and Attitudinal Market Segmentation. *Transport Policy*, Vol. 15, 2008, pp. 186–195.
9. Outwater, M. L., S. Castleberry, Y. Shiftan, M. Ben-Akiva, Y. S. Zhou, and A. Kuppam. Attitudinal Market Segmentation Approach to Mode Choice and Ridership Forecasting: Structural Equation Modeling. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1854, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 32–42.
10. Anable, J. “Complacent Car Addicts” or “Aspiring Environmentalists”? Identifying Travel Behaviour Segments Using Attitude Theory. *Transport Policy*, Vol. 12, 2005, pp. 65–78.
11. Beirao, G., and J. S. Cabral. Market Segmentation Analysis Using Attitudes Toward Transportation: Exploring the Differences Between Men and Women. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2067, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 56–64.
12. Krizek, K. J. Lifestyles, Residential Location Decisions, and Pedestrian and Transit Activity. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1981, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 171–178.
13. Pas, E. I., and J. C. Huber. Market Segmentation Analysis of Potential Inter-City Rail Travelers. *Transportation*, Vol. 19, 1992, pp. 177–196.
14. Dallen, J. Sustainable Transport, Market Segmentation and Tourism: The Looe Valley Branch Line Railway, Cornwall, UK. *Journal of Sustainable Tourism*, Vol. 15, No. 2, 2007, pp. 180–199.
15. Dallen, J. The Challenges of Diverse Visitor Perceptions: Rail Policy and Sustainable Transport at the Resort Destination. *Journal of Transport Geography*, Vol. 15, 2007, pp. 104–115.
16. Outwater, M., K. Tierney, M. Bradley, E. Sall, A. Kuppam, and V. Modugula. California Statewide Model for High-Speed Rail. *Journal of Choice Modelling*, Vol. 3, No. 1, 2008, pp. 58–83.
17. Transportation Economics & Management Systems, Inc., and HNTB Corporation. *Midwest Regional Rail System*. Executive summary report. Midwest Regional Rail Initiative, 2004 <http://www.dot.state.wi.us/projects/state/docs/railmidwest.pdf>. Accessed July 23, 2010.
18. Amtrak System Timetable: Fall 2009–Winter 2010. National Railroad Passenger Corporation, Washington, D.C., 2009.
19. Sperry, B., and E. Johnson. Examining the Role of the *Hiawatha Service* in the Milwaukee–Chicago Corridor. *Proc., 2010 Joint Rail Conference*, Urbana, Ill., 2010.

---

*The views expressed in this paper are the opinions of the authors and do not represent official views of the Texas Transportation Institute, the Wisconsin Department of Transportation, or Amtrak. Any errors, inaccuracies, or omissions are the responsibility of the authors.*

*The Intercity Passenger Rail Committee peer-reviewed this paper.*

# High-Speed Railways in Spain

## Example of Success?

M. Sánchez-Borràs, F. Robusté, and O. Criado

The Spanish high-speed railway (HSR) network has developed swiftly over the past few years. In less than two decades, this network has already become the second longest in Europe and the third in the world ranking, and the network continues to grow. Given this situation and given that the Spanish HSR system is considered by several governments as an example to follow, the authors of this paper analyzed whether the Spanish case could be considered an example of success in terms of HSR and whether some Spanish practices could be improved on. A thorough analysis presents the Spanish HSR network in terms of its use, competitiveness with air travel, environmental and territorial impacts, and profitability and costs. The results of this analysis allowed conclusions to be drawn regarding the success of the Spanish HSR and some recommendations to be made for decision makers planning the construction of a high-speed line.

The first analyses of the viability of a high-speed line (HSL) in Spain to link Madrid with Barcelona, Spain, and France were undertaken in the 1970s. However, the idea did not prosper because, according to the results of the profitability study, there would not be sufficient traffic to justify a new line until 2000.

In 1979, the general railway master plan was passed. Although it did not consider the construction of an HSL, this plan did envisage the construction of three new lines to solve the capacity problems Spanish rail was experiencing. In 1986, the decision was made to construct one of these lines, the New Railway Access to Andalusia, which was intended to solve the limited-capacity problems of the Getafe–Córdoba section between Madrid and Seville, Spain. Construction began in 1988 with some strategic changes to the project that would give rise to the first HSL in Spain. Specifically, the decision was made to build the new line using the international gage (1,435 mm) instead of the Iberian one (1,668 mm) and to extend the line directly to Seville. In addition, the Seville–Madrid–Barcelona–French border axis (with the international gage) was approved. Five years later, the first Spanish HSL (Madrid–Seville) was opened to traffic.

In 2004, the Spanish government approved the master plan on multimodal transport infrastructure and services. This plan represents an important milestone in the definition of the transport system, the promotion of railways, and above all, the important decision to opt for high-speed railways (HSRs). The terms of this plan are important for the following reasons:

- It promotes the construction of an HSR network that in 2020 should reach 7,200 km (4,472 mi; 1 mi = 1.61 km);
- It contributes to the accessibility by rail of the entire country, aiming at placing 90% of the Spanish population within 50 km of an HSR station; and
- It plans to offer travel times by HSR that are competitive with air travel for travel distances less than 700 km and with private vehicles for distances greater than 300 km.

Fifty percent of the master plan's budget is devoted to railways, that is, about €83,000 million ( $\approx$  \$107,900 million; €1  $\approx$  \$1.3 in November 2010 U.S. dollars). To help prevent Spain's current economic crisis from hindering the development of HSR, in April 2010 the Spanish government approved an extraordinary infrastructure plan promoting railways in an effort to hasten economic recovery by creating new jobs.

The implementation of the master plan on multimodal transport infrastructure and services has resulted in an average yearly investment in the railway sector of about €6,000 million and in the opening of 580 km of new HSL from 2005 to 2008. In September 2009, with 1,599 km of HSL in operation (Figure 1a), Spain was third in the top world ranking, after Japan (2,145 km) and France (1,872 km), and had the world's longest HSR corridor (Barcelona–Málaga, 1,121 km).

The current network places 40% of the Spanish population within 50 km of an HSR station (Figure 1b and 1c). In addition to these 1,599 km, 2,200 km are under construction and 1,700 km more are planned, making a total of 5,520 km (almost one-third of the length planned in all Europe).

The Spanish HSR network has developed very quickly. In less than two decades, it is almost as long as the French network, even though the first HSL in France was opened in 1981 (1). Once the lines under construction are opened to traffic, Spain, with 46,600,000 inhabitants and 506,990 km<sup>2</sup>, will be the country with the most HSL after China (1,300,000,000 inhabitants and 9,571,300 km<sup>2</sup>). Given this rapid growth, it is of great interest to analyze whether the Spanish case can be considered as an example of HSR success and whether some Spanish practices may be improved on.

The aim of the present paper is to thoroughly analyze the Spanish HSR network in terms of its use, competitiveness against airways, environmental and territorial impacts, and profitability and costs in order to come to some conclusions on its success and give some recommendations to bear in mind when planning the construction of an HSL.

## CHARACTERIZATION OF USE OF SPANISH HSR NETWORK

### Supply in Spanish HSR Network

Spanish HSLs are currently used only for passenger traffic. Even though mixed traffic use has been projected for some of the lines,

M. Sánchez-Borràs and O. Criado, Center for Innovation in Transport, Jordi Girona 29, 2-A, Building Nexus II, and F. Robusté, Jordi Girona 1-3, Building B1, Room 211, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain. Corresponding author: M. Sánchez-Borràs, marta.sanchez-borras@upc.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 39–48.  
DOI: 10.3141/2261-05

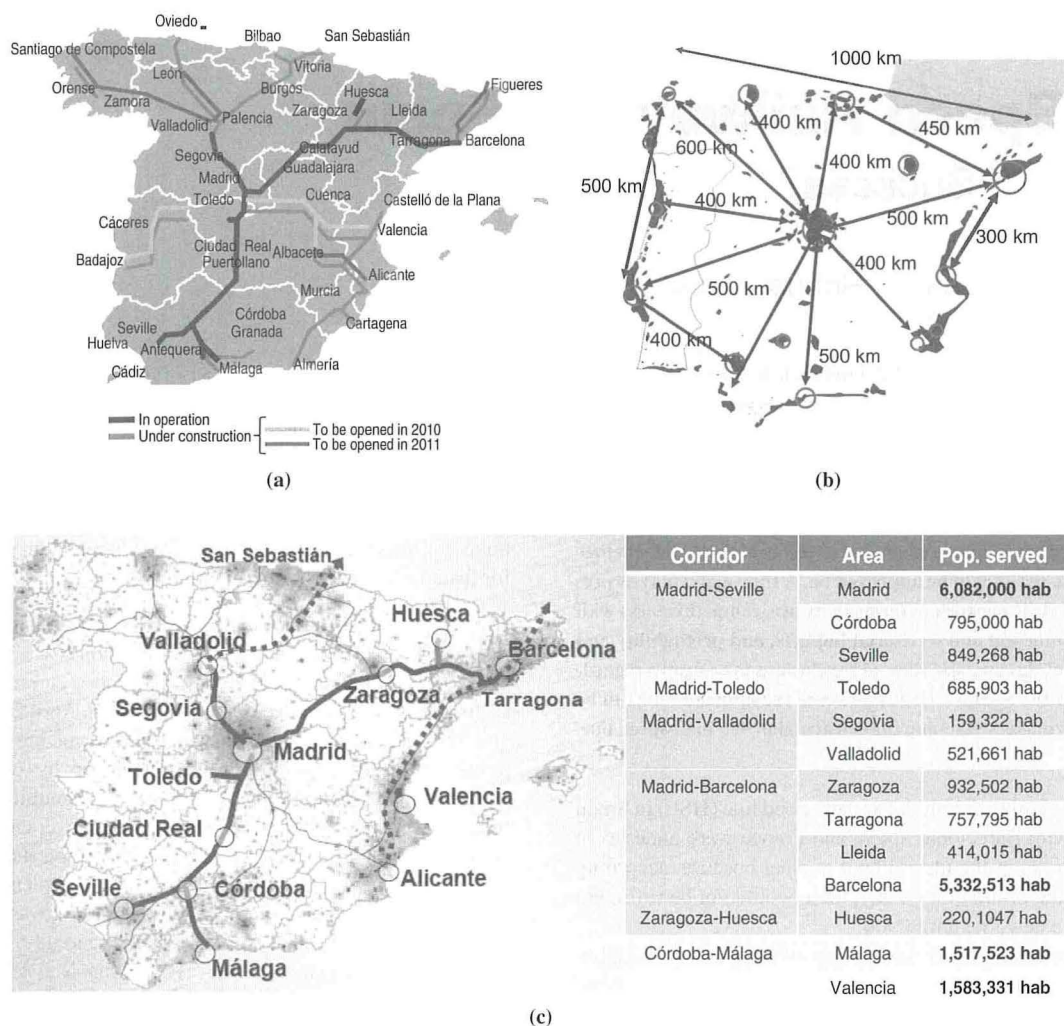


FIGURE 1 Spanish HSR network: (a) lines in operation and under construction, (b) distances between cities, and (c) population served (pop. = population, hab = inhabitants).

this option is not likely given the advantages of a one-use-only rail network compared with a mixed-use one and because conventional line capacity has increased due to the shift of passenger trains from conventional lines to HSL.

The railway passenger services running on HSL can be grouped into long-distance (commercial) and medium-distance (public) services. Long-distance services can be divided into two types: those running exclusively on HSL (built with the international gage) and those running on both HSL and conventional lines (built with the Iberian gage) using variable-gage trains. AVE trains are found in the first group, and Alvia, Altaria, and Talgo trains are in the second group.

Figure 2 and Table 1 summarize the main characteristics of the long-distance HSR services. AVE services link distances of 475 to 630 km in approximately 2 h 30 min, thanks to commercial speeds ranging from 200 to 240 km/h for nonstop trains (the only exception is the Madrid–Valladolid HSL, which constitutes the first section of the Madrid–Northeast Spain HSL). Three ticket fares (economy, preferential, and club) have been introduced and set according to the commercial speed offered (the faster the train, the higher the fare), the type of vehicle, the onboard service provided, and the level of rail charges to be paid by the railway undertaking to the infrastructure manager. Fares range from 0.17 to 0.21 (euro) cents/km for the economy class to 0.30 to 0.38 (euro) cents/km for the first class (club).

Alvia, Altaria, and Talgo services offer commercial speeds ranging from 130 to 189 km/h and fare rates similar to those of AVE services with the difference that only two fare rates are available (economy and preferential).

Spain is the world pioneer in the provision of regional HSR services (Avant), which were introduced to link municipalities with high mobility needs that are less than 200 km apart. These regional commercial HSR services are characterized by travel times below 1 h 30 min, maximum speeds of 250 km/h and commercial speeds around 150 km/h, high frequencies, adapted rolling stock with smaller coaches, and specific timetables adapted to the needs of the users.

Because most regional HSR users are price-sensitive commuters, a differentiated single-pricing structure adapted to the commuters' profile has been implemented. Since these services are considered public services, their pricing structure benefits from public subsidies. These subsidies result in fares ranging from 42% to 57% less than the full price of the economy fare for nondirect long-distance HSR trains (Table 2). The subsidies for the public service fares are negotiated on a multiyear basis. Round-trip tickets as well as rail cards with 20 to 50 trips and special discount tickets are available.

The cities covered by regional HSR services (Figure 2) are linked with travel times ranging from 30 min to 1 h 10 min for distances of 40 to 200 km (Table 2). Service frequencies vary from

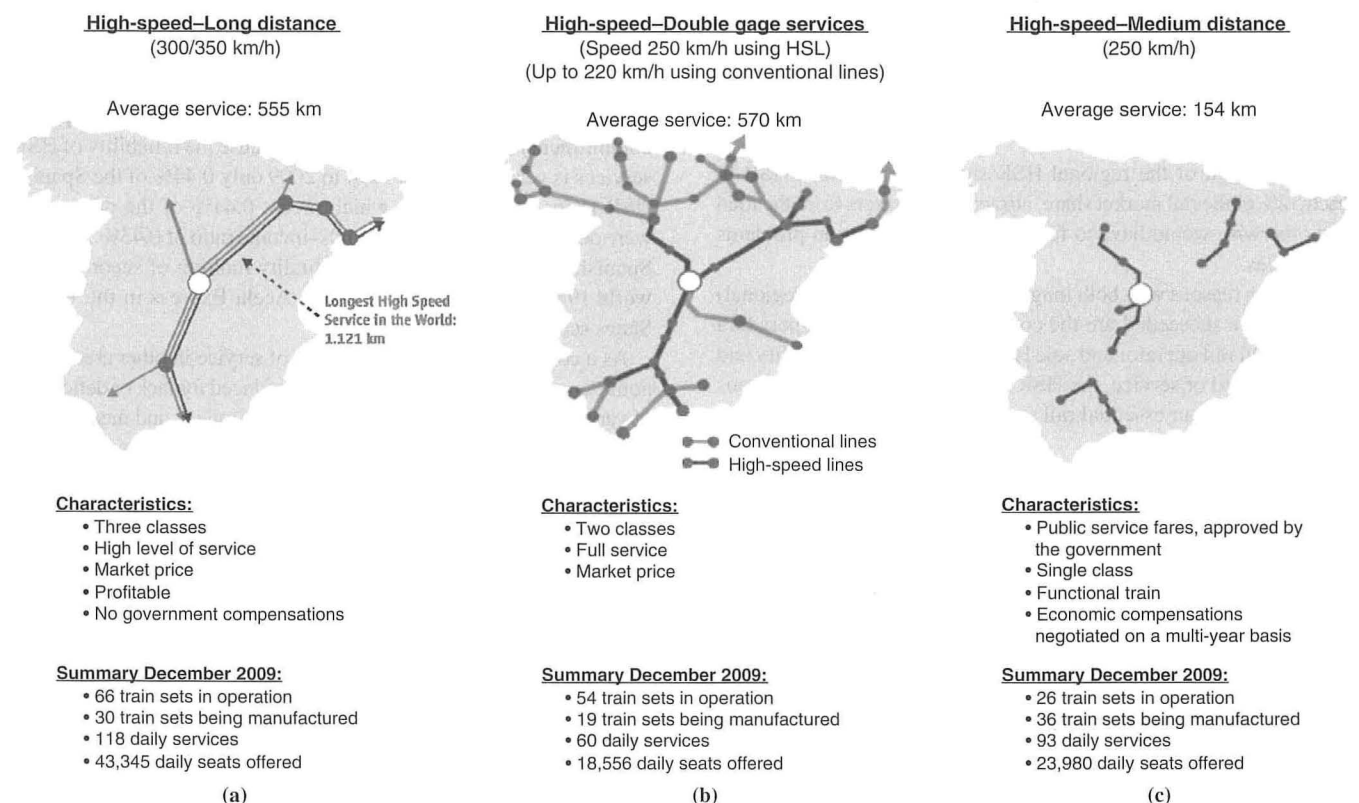


FIGURE 2 Spanish HSR services: (a) AVE, (b) Alvia, and (c) Avant.

TABLE 1 Long-Distance HSR Services

		Fare [(euro) cent/km]									
		$V_{com}$ (km/h)		Travel Time		Economy		Preferential		Club	
AVE Services	L (km)	NS	S	NS	S	NS	S	NS	S	NS	S
Madrid-Barcelona	629	239	191	2 h 38 min	3 h 18 min	0.2138	0.1811	0.3207	0.2719	0.3847	0.3261
Madrid-Seville	473	203	183	2 h 20 min	2 h 35 min	0.1706	0.1535	0.2558	0.2302	0.3070	0.2763
Madrid-Málaga	500	200	176	2 h 30 min	2 h 50 min	0.1700	0.1528	0.2548	0.2294	0.3058	0.2754
Madrid-Valladolid	178	191	NA	56 min	NA	0.1972	0.1972	0.2612	0.2612	NA	NA

NOTE: L = link distance;  $V_{com}$  = commercial speed; NS = nonstop services; S = services with stops; NA = not applicable. Alvia, Altaria, and Talgo services have an average distance of 570 km, speeds of 130–189 km/h, and economy and preferential rates similar to AVE services.

TABLE 2 HSR Regional Services

HSR Regional Service	Travel Time	L (km)	$V_{com}$ (km/h)	Service Frequency (trains/day/direction)	Avant Full-Price Fare (€/km)	AVE Economy Fare (€/km)
Barcelona-Tarragona-Lleida	1 h 10 min	169	145	6	0.1331	0.2692
Madrid-Segovia	30 min	67	134	3	0.1478	0.3418
Madrid-Toledo	30 min	38	76	11	0.2605	NA
Madrid-Ciudad Real-Puertollano	1 h 13 min	200	164	12	0.1285	0.2195
Córdoba-Seville	45 min	127	169	8	0.1256	0.2276
Córdoba-Málaga	1 h 5 min	154	142	6	0.1364	0.2883

link to link, with Madrid–Ciudad Real (171 km) having the highest frequencies. This service was the first to be implemented (in 1994) and since then, Ciudad Real has grown considerably as a result of being able to attract a new population that can make a daily commute to Madrid. A similar impact is expected for the Barcelona–Tarragona–Lleida and the Madrid–Segovia links, in operation since 2008.

Development of the regional HSR supply has allowed HSR to reach 52% of the rail market share, attracting new users from the road who otherwise would have to face significant congestion problems in big cities.

The main reasons why both long- and medium-distance (regional) services have succeeded are the commitment by Renfe Operadora (the national rail operator and sole HSR operator) to punctuality and a high standard of service, the HSR competitive pricing conditions (Figure 3), and four essential rail services attributes:

- Safety. To date there have been no accidents with deaths on newly built HSLs anywhere in the world.
- Comfort. HSR trains offer bigger seat pitches than airplanes.
- Competitive journey times. HSR services equal door-to-door times offered by airways.
- Environmental friendliness. Rail transport has less environmental impact than its competitor modes (see section on HSR impacts).

The quality commitment of Renfe Operadora, intended to attract more clients, is a pioneer in the rail sector and emphasizes punctuality and a high standard of service. The punctuality commitment was first introduced in 1994 in HSR services. It consisted of refunding (in cash) the total ticket price if a train arrived at its destination more than 5 min late. The success of the on-time commitment on AVE trains led to the gradual implementation of this commitment in other long- and medium-distance services. Renfe Operadora's cur-

rent delay refunds are summarized and compared with those offered by other European railway operators in Table 3.

The comparison shows that Renfe Operadora's commitment to HSR timeliness is by far the most competitive, both in terms of delay and type of refund. Spain is the only country where refunds are paid in cash (together with Germany for Thalys services). This quality commitment is financially acceptable because the reliability of HSR services is very high: from 1994 to 2009 only 0.44% of the Spanish HSR trains were delayed, which meant 0.41% of the passengers were delayed, resulting in a refunds–income ratio of 0.43%. In 2008, Spanish HSR services had a punctuality ranking of second in the world (98.5%) after Japan (99.0%). Acela Express in the United States scored 77.0%.

As a commitment to high standards of service in other areas, economic compensations have also been introduced for lack or deficiency of various services such as air-conditioning, toilets, and meal service at the seat and in restaurants or cafeterias. Compensations range from 15% to 100%, depending on the importance given to each service.

The commitment to quality has increased clients' trust and satisfaction. In 2009, only 1% and 2% of the Madrid–Seville and Madrid–Barcelona HSL users, respectively, declared themselves dissatisfied with the service received. This high customer satisfaction places HSR in a competitive position against air services.

### Demand in Spanish HSR Network

Figure 4 presents the demand on the Spanish HSR network and in different HSR corridors. In 2008, demand doubled as a result of the opening of the Madrid–Barcelona HSL. Before 2008, the Madrid–Barcelona link carried approximately 750,000 passengers per year; after the opening of the HSL in 2008, the demand more than tripled, reaching 2.75 million passengers in 2009.

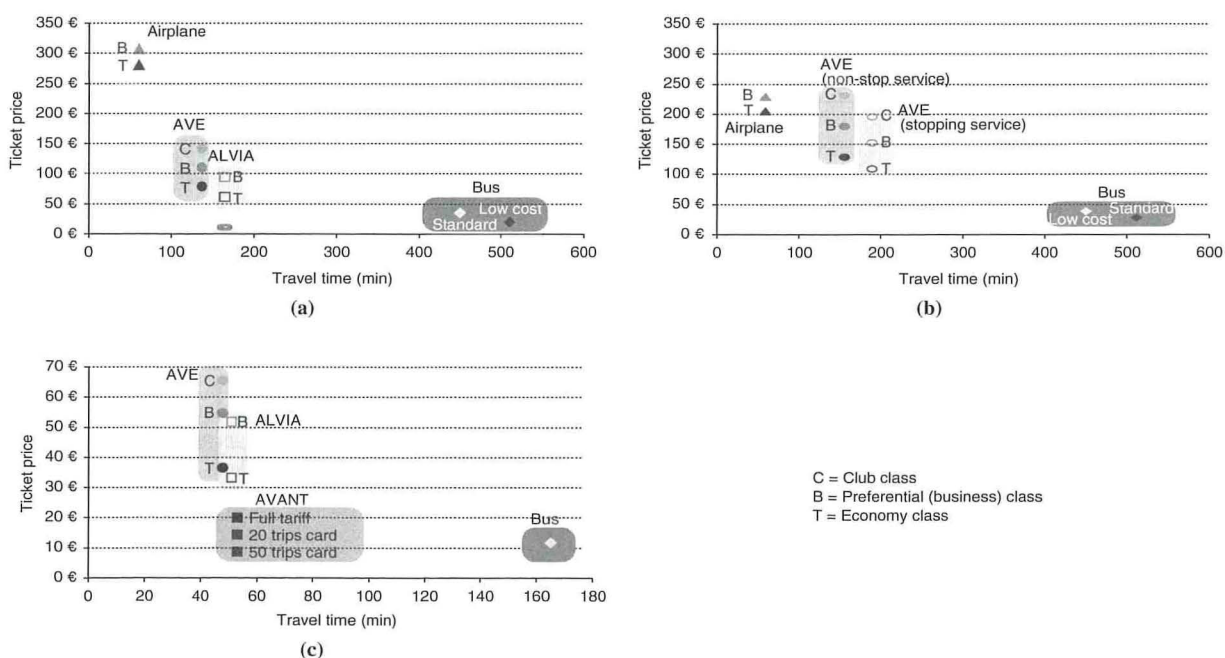


FIGURE 3 HSR pricing conditions compared with those of competitors: (a) Madrid–Seville, 471 km; (b) Madrid–Barcelona, 628 km; and (c) Madrid–Ciudad Real, 171 km.

TABLE 3 Compensation for Delays Offered by European Rail Operators

Country (Operator)	Products or Type of Service	Delay	Refund of Ticket Price
Spain (Renfe Operadora)	Madrid–Seville AVE trains	>5 min	100%
	Madrid–Puertollano Avant trains	>5 min	100%
	Córdoba–Seville Avant trains	>5 min	100%
	Any other AVE or Avant trains (temporarily)	>15 min	50%
		>30 min	100%
	Alvia, Alaris, Altaria, Euromed trains	>20 min	25%
		>40 min	50%
		>60 min	100%
	Regional trains	15–29 min	20%
		30–59 min	50%
		>60 min	100%
France (SNCF)	TGV: Journeys of more than 100 km	>30 min	One-third <sup>a</sup>
Germany (DB)	Main-line trains: ICE, IC, EC, IR, MET, and Thalys, and CIS on national routes	>60 min (including missed connection)	20% <sup>a</sup>
	IC-Sprinter	>30 min	Value of IC-Sprinter supplements <sup>a</sup>
	DB NZ, UEx, CNL, EN, D	>120 min (including missed connection)	20% of outbound price (including additional prices) <sup>a</sup>
Italy (FS)	Italian Eurostar trains	>25 min	50% <sup>a</sup>
	IC and EC trains on national routes	>30 min	30% of ticket price and used reservation <sup>a</sup>
	Overnight IC trains and express trains	>60 min	30% of ticket price and used reservation or seat and 20% of couchette or bed service price <sup>a</sup>
Portugal (CP)	Quality titling trains on main-line and IC conventional trains	60–90 min	50%
		>90 min	100%
France, Germany, Belgium, Netherlands	Thalys (international journeys only)	>30 min	20% in vouchers (cash in Germany)
		>60 min	50% in vouchers (cash in Germany)
		>120 min	100% in vouchers (cash in Germany)
United Kingdom, France	Eurostar (all services, except Avignon, France, and ski trains)	60–180 min	50% <sup>a</sup> or free outbound journey
		>180 min	Cash refund

NOTE: SNCF = Société Nationale des Chemins de Fer Français; DB = Deutsche Bahn; FS = Ferrovie dello Stato; CP = Comboios de Portugal; ICE = intercity express; IC = intercity; EC = Eurocity; IR = interregional; MET = metropolitan; NZ = night train; UEx = Urlaub Express; CNL = City Night Line; EN = EuroNight.

<sup>a</sup>Refund in travel vouchers.

SOURCE: Renfe Operadora and Charter on Rail Passenger Services in Europe.

Similarly, the Madrid–Seville HSR link, which was opened in 1992, has shown a steadily increasing demand because of some network effects. It is also relevant to mention the Madrid–Toledo link, which is used by HSR regional services and attracts 1.5 million passengers per year, and the Madrid–Valladolid and the Camp de Tarragona–Madrid links, which both are (or will be) intermediate stops of a longer HSL. In the case of Camp de Tarragona, the demand headed toward Madrid is very low (around 300,000 passengers per year), but the one toward Barcelona served by HSR regional trains has a special interest for potential commuters.

### COMPETITIVENESS: MARKET SHARE OF SPANISH HSR AND AIRWAYS

The increase in rail demand has translated into an increase in rail market share. Figure 5 summarizes the market share evolution for the main Spanish HSR links.

In the case of the Madrid–Barcelona link, HSR offers a commercial speed of 240 km/h, allowing rail travel time from city center to city center to drop from 7 h to 2 h 38 min. As a result of this time reduction, the rail–air market share increased from 12% to 48% for railways.

In the case of the Madrid–Seville link, a commercial speed of 200 km/h was high enough to reach a competitive travel time with airways. The opening of this HSR link meant an increase of the total rail–air market share of approximately 60 points (from 21% to

82%). In the global mode split, automobile travel lost half its initial market share.

The data presented show that the consolidation of Spanish HSR with respect to air transport has resulted in a new modal distribution in which HSR dominates.

### IMPACTS PRODUCED BY SPANISH HSR

#### Environmental Impacts

Figure 6 presents the results of a study in which García-Álvarez compared the carbon dioxide (CO<sub>2</sub>) emissions per passenger kilometer for five transport modes (2). Calculations were made for three links served by HSR services, taking into consideration the assumptions on load factor and vehicles (Figure 6). The results obtained show that for the links analyzed, railways emit five times less CO<sub>2</sub> per passenger kilometer than automobiles, six times less than airways, and approximately the same amount as bus services for long-distance links and conventional rail. (For regional services such as Madrid–Toledo, bus is the transport mode that emits less CO<sub>2</sub> per passenger kilometer.) Pérez-Martínez and Monzón report results of the same order of magnitude (3). It can therefore be concluded that the increase in HSR market share has had a positive environmental impact in terms of reduced greenhouse gas emissions.

Environmental impacts have been positive in terms of energy consumption as well. Figure 6 shows that HSR consumes less energy than

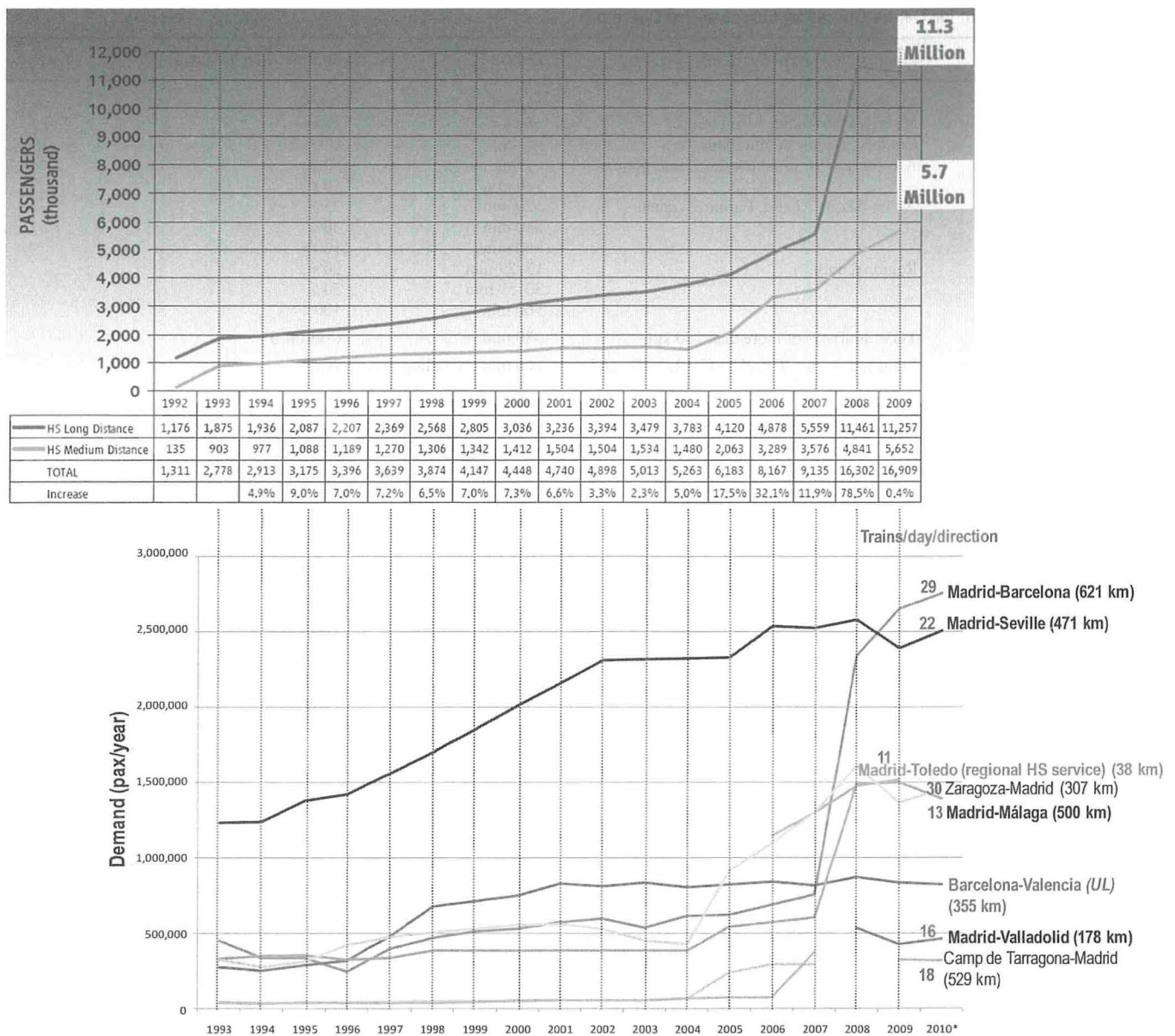


FIGURE 4 Demand on Spanish HSR network (pax = passenger; UL = upgraded line; \* = data are estimated on the basis of first 6 months of 2010). (Source: Developed from Renfe Operadora data.)

its competitor modes and approximately the same as conventional rail. The reasons for this are given by García-Álvarez (2).

### Territorial Impacts

Impacts have not only been environmental, but territorial as well. Temporal distance has replaced spatial distance since today traveling 170 km by HSR takes as long as a 70-km trip using conventional services. This change of paradigm has increased commuting mobility.

In addition, the boost given to Spanish HSR has had a great economic impact. The more than €860 million exported by the railway sector in 2009 and the creation of 116,000 jobs support this statement. Furthermore, increased HSR has led to important urban transformations. For example, Ciudad Real has grown considerably and broadened its influence areas as a result of the introduction of regional HSR services.

### PROFITABILITY AND COSTS OF SPANISH HSR

It is well-known that HSR requires important investments that have to be recovered. In the Spanish case, those investments are around €15 to €24 million per kilometer. In the current European railway scenario, rail charges paid by railway undertaking to the infrastructure manager for using its tracks are the main source of income for the infrastructure manager, and consequently the system's viability depends on the capacity of HSR to attract clients. Therefore, the use of HSL should be maximized. That is, new strategies should be established to attract traffic, incentives for railway undertaking and final clients should be generated, and old and new markets should be attracted to the railways. One of the main tools for achieving maximization of HSL use is the structure and level of rail infrastructure charges.

According to European Directive 2001/14/EC, rail charges should consist of a basic charge set at the marginal cost (MC) of infrastructure (in the railway field, MC corresponds to the variation of infra-

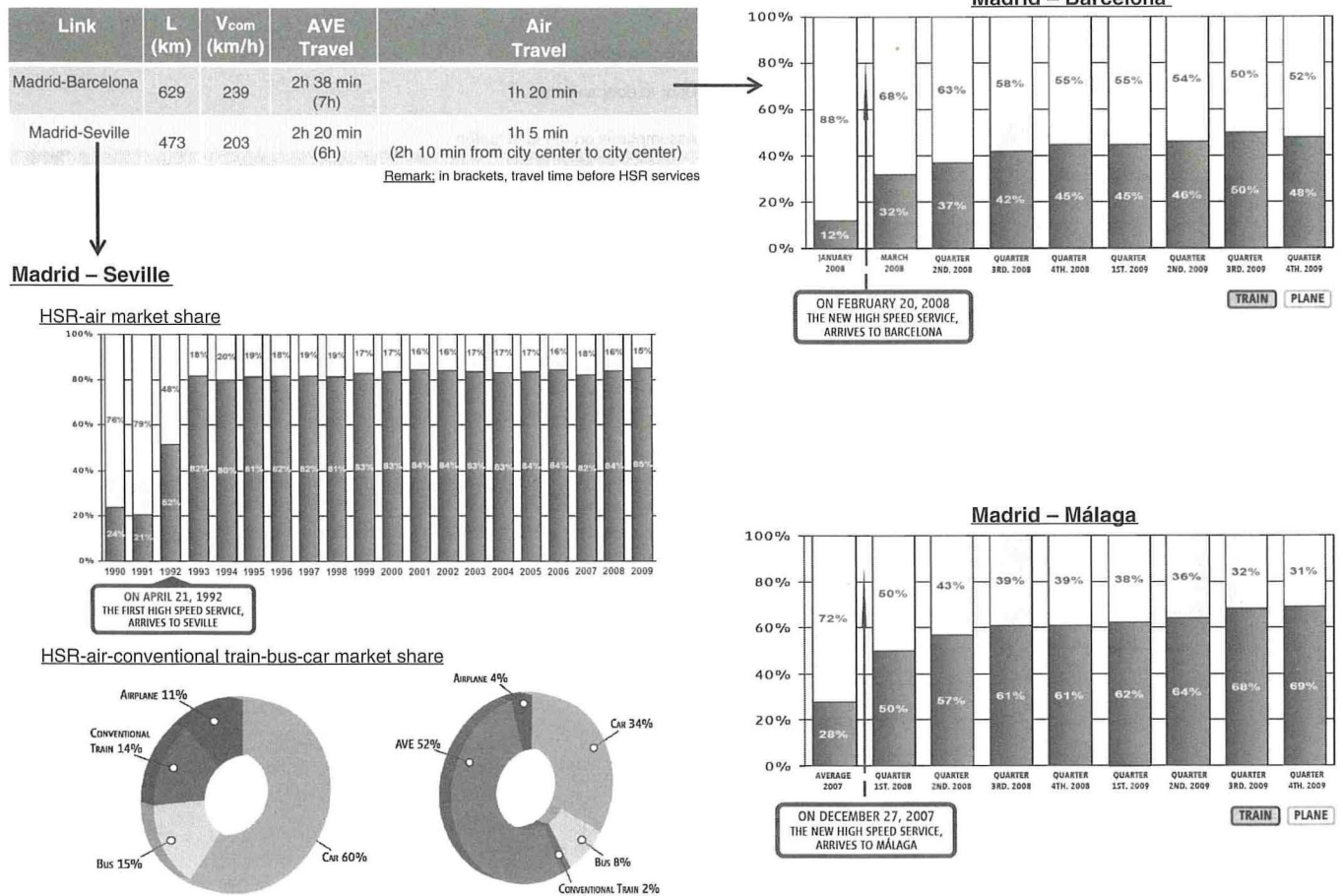


FIGURE 5 Yearly and monthly market share evolution.

structure costs when a unit of additional traffic uses the tracks). The basic charge may also include a charge to internalize external costs if this is practiced by the modes competing with railways; a scarcity charge for specific sections at specific time periods; and some markups above MC, aimed at recovering costs, when the market can bear them.

Compared with the charging structure recommended by the European legislation, the Spanish charging structure (4, 5) does not clearly state how markups have been set. In any case, markups above MC of infrastructure aimed at recovering (part of) the fixed costs should not challenge the economic viability of railway undertaking and should minimize the number of passengers deterred from using HSR as a result of too high a charge.

Table 4 shows that current rail charges applied in the Spanish HSR network do not challenge the operator's economic viability (the demand generated allows the operator to cover operating costs). With regard to the second point, the optimal charge can be calculated with the formula presented by Sánchez-Borràs (6) and deduced by Crozet (7) from the Ramsey pricing formula (8, 9), that is, with Equation 1:

$$a = \frac{C_i + \frac{\alpha}{\epsilon} \cdot C_s}{1 - \frac{\alpha}{\epsilon}} \quad (1)$$

where

$a$  = level of infrastructure charge;  
 $C_i$  = marginal infrastructure cost;

$\alpha = \lambda/(1 + \lambda)$ , a parameter reflecting the opportunity cost of public funds  $\lambda$ ;

$\epsilon$  = price elasticity of traffic expressed in absolute value;

$C_s$  = marginal train service cost; and

$C_s + C_i = C$ , which is the overall MC.

The application of this formula to the Spanish case gives an optimal markup above MC for HSR services of around €4 per train kilometer at normal hours, when elasticities are higher, and €8 per train kilometer at peak periods, when elasticities are lower. From these values it may be concluded that the current charges applied to Spanish HSR links are higher than the optimal ones (Figure 7).

To quantify what would happen if current infrastructure charges were reduced to the optimal level, the formula defined by Sánchez-Borràs (6), which allows determining the relative change in traffic resulting from a reduction in rail charges, was used:

$$\frac{\Delta Q}{Q} = \frac{\epsilon}{1 + \epsilon} \cdot \gamma \cdot \frac{\Delta IC}{IC} \quad (2)$$

where

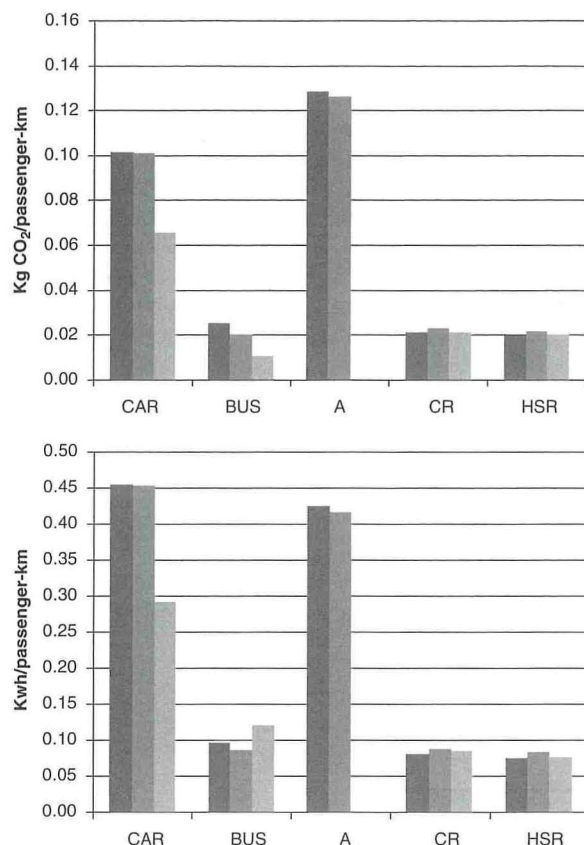
$Q$  = traffic expressed in number of passengers per train,

$\epsilon$  = price elasticity of traffic expressed in absolute value,

$\gamma$  = ratio between infrastructure charges and revenues from ticket sales for a train running on a given link, and

$IC$  = rail charges.

■ MADRID-SEVILLE ■ MADRID-BARCELONA ■ MADRID-TOLEDO



#### Calculation assumptions:

a) Door to door services

b) Assumptions on the load factor:

Transport mode	Madrid-Seville	Madrid-Barcelona	Madrid-Toledo
Car	0.3	0.3	0.4
Bus	0.55	0.61	0.52
Airplane (A)	0.75	0.75	
Conventional rail (CR)	0.64	0.64	0.35
HSR	0.7	0.7	0.56
Cars for access to airports and stations	0.3	0.3	0.3

c) Assumptions on vehicles:

Transport mode	Type of vehicle
Car	Long distance: petrol mid-size car and diesel oil small family car Madrid-Toledo: small diesel oil car and small family diesel oil car
Bus	Madrid-Barcelona and Madrid-Seville: 56.26 seats Madrid-Toledo: 50 seats
Airplane (A)	170 seats (15% airplanes are MD80-old-and 85% A320)
Conventional rail (CR)	Trains used for day services before AVE: 282 seats for M-S; 228 seats for M-B; 270 seats for M-T
HSR	Madrid-Seville: series 100 and 103; 367 seats Madrid-Barcelona: series 102 and 103; 351 seats Madrid-Toledo: series 104.1; 270 seats
Cars for access to airports and stations	Diesel oil mid-size car

FIGURE 6 CO<sub>2</sub> emissions (top) and energy consumption (bottom left) per passenger kilometer in three HSR links. [Source: Developed from García-Álvarez (2).]

For the calculations it was assumed that the impact on traffic volumes would be calculated for the peak period. The assumptions for the calculation of infrastructure charges and revenues from ticket sales are presented elsewhere (6).

With these assumptions it can be affirmed that if rail charges were reduced to bring them to the optimal level, the number of HSR passengers would increase by 6.19% to 14.90% during the peak period (47,086 to 163,675 passengers per year) for price elasticities of demand of 0.57 and by 15.46% to 40.18% in the peak period (99,538 to 441,338 passengers per year) for price elasticities of demand of

0.7 (Table 5). In other words, an important number of passengers are deterred from using HSR as a result of applying too high a charge.

## CONCLUSIONS

Four main conclusions can be drawn from the information presented in the preceding sections:

First, Spain's decision to advance HSR has allowed HSR to succeed in competition with airways. This success can be explained

TABLE 4 Profitability of Spanish HSR Services

Income or Expense	Madrid-Barcelona (nonstop)	Madrid-Zaragoza-Barcelona	Madrid-Valladolid	Madrid-Málaga	Madrid-Seville
Railway undertaking income (€ thousands)	35,298	30,679	7,199	18,443	17,371
Railway undertaking expenses (€ thousands)	16,994	16,703	5,225	12,435	11,730
Income - expenses (€ thousands)	18,304	13,977	1,975	6,008	5,641
Rail charges (€ thousands)	6,126	6,044	1,345	4,068	3,859
Benefits (€ thousands)	12,178	7,933	630	1,940	1,781
Rail charges/income (%)	17	20	19	22	22
Benefits/income (%)	35	26	9	11	10

SOURCE: Administrador de Infraestructuras Ferroviarias.

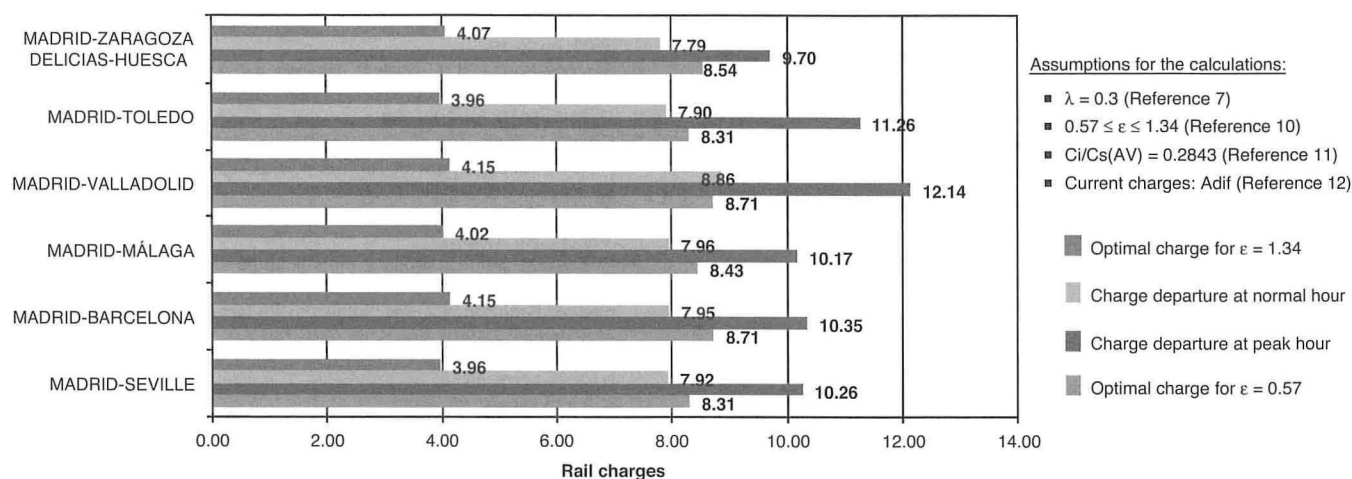


FIGURE 7 Current versus optimal charges (7, 10–12).

mainly by the increase in commercial speed, which allows HSR to reach door-to-door times compatible with those offered by airways. For instance, in the Madrid–Barcelona HSL it was necessary to reach 350 km/h in order to reduce the travel time to 2 h 30 min for HSR to cover two-thirds of the rail–air market share. If the travel time had not been reduced by that much, air travel would have retained two-thirds of the market share. Currently, the rail–air market share is approximately fifty–fifty because of the technical difficulties involved in reaching a maximum speed of 350 km/h on a ballasted track.

The Spanish HSR success can also be explained by the high-frequency service it offers, its high safety and reliability records, and the importance given by the Spanish population to environmental sustainability, which should favor HSR even more once the external costs pricing policy is implemented in all transport modes.

Second, while operative decisions take economic rationality into account by considering the results of cost–benefit analysis, strategic decisions allow taking some hidden attributes other than time costs into account (e.g., comfort and reliability for the users) that transportation projects such as HSR may provide. Among those hidden attributes are the importance of strengthening territorial links, the increase of acceptable commuting distance to more than 100 km, and the economic growth that the rail industry can support. Therefore, decisions regarding HSR network implementation should consider all the effects of the projects for all the stakeholders, and input–output tables and multicriteria analyses such as those described in project appraisal guidelines for railways [e.g., *Railway Project Appraisal Guidelines* (13)] should complement cost–benefit analysis.

TABLE 5 Impact on Traffic Volumes from Reduction of Rail Infrastructure Charges to Optimal Level

Link	$\Delta Q/Q$		$\Delta Q$ (passengers/year)	
	$\epsilon = 0.57$ (%)	$\epsilon = 0.7$ (%)	$\epsilon = 0.57$	$\epsilon = 0.7$
Madrid–Seville	6.19	20.60	47,086	156,660
Madrid–Barcelona	4.15	15.46	35,092	130,673
Madrid–Málaga	5.54	19.63	28,073	99,538
Madrid–Valladolid	14.90	40.18	163,675	441,338
Madrid–Toledo	6.16	17.23	52,092	145,606

Spain has made a strategic decision, but perhaps the trade-off between spatial coverage and critical mass of demand has not been considered thoroughly. While in the Madrid–Barcelona HSL the 350 km/h speed is justified, in other cases 200 km/h is enough to attract all the potential demand. Therefore, when considering whether to build an HSL or to upgrade an existing line (a considerably cheaper option than building an HSL) to reach speeds of 200 km/h, one should bear in mind that time reductions reached through speed increases should compensate increases in social and financial costs and, therefore, the price elasticity of demand should be evaluated.

In Spain, the government planned the HSR network under the slogan “no provincial capital at more than 3 h from the country’s capital,” but taking into account the considerations discussed, it would seem more reasonable to have used the slogan “No provincial capital of more than X inhabitants at more than Y minutes from an HSR station.”

Third, it can be concluded that it is necessary to justify the economic profitability and financial equilibrium of all HSL investments. Without this justification rail operators can end up with a very expensive low-demand HSL, such as the Madrid–Valladolid line, which is part of the HSL that will link Madrid to Northeast Spain. According to forecasts, the demand generated with the entire line in operation should cover operating and infrastructure costs. However, this HSL link currently ends in Valladolid, which has less population than Northeast Spain (Figure 1b and 1c). The lower demand (approximately 500,000 passengers per year) does not cover total infrastructure manager expenses, mainly infrastructure costs (€26.3 million in 2008), and therefore incurs losses (€18.8 million in 2008). In such cases, the low demand translates into low incomes from rail charges. To overcome this, it would be necessary to raise infrastructure charges to ensure recovery of the costs of operation activities. However, this would cause a problem, because such an increase of charges (the increase of markups above MC), would result in a traffic reduction.

Finally, it can be affirmed that there is a need to redefine the rail charges structure and level. As the European Conference of Ministers of Transport declared in 1998, rail charges are a “tool for modernizing railways organization and rendering it more efficient, i.e., more able to face competition of other modes of transport” (14). Therefore, to render HSR more able to face the competition of other modes of transport, rail charges should clearly reflect both infrastructure marginal and external costs, and markups above MC should be set according to the market’s ability to pay in order to maximize

social benefits. This implies eliminating the fixed access charge (which is discriminatory in an open market scenario and therefore not consistent with Directive 2001/14/EC) and estimating the price elasticity of demand, the budget restriction, and the marginal infrastructure and service cost for each link in order to obtain the optimal markup for each link. With this information, the connections between all the stakeholders and all the effects should be analyzed to ensure the maximization of social benefits.

## REFERENCES

1. López-Pita, A., P. F. Teixeira, C. Casas, and A. Bachiller. Practical Experience on the Commercial Impact of High Speed Rail Services on the Competing Modes in Europe: 1981–2006. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
2. García-Álvarez, A. Consumo de energía y emisiones del tren de alta velocidad en comparación con otros modos de transporte. *Anales de Mecánica y Electricidad*, 2007, pp. 26–34.
3. Pérez-Martínez, P. J., and A. Monzón. Consumo de energía por el transporte en España y tendencias de emisión. *Observatorio Medioambiental*, Vol. 11, 2008, pp. 127–147.
4. Sánchez-Borràs, M., and A. López-Pita. Rail Infrastructure Charging Systems for High-Speed Lines in Europe. *Transport Reviews*, October 20, 2010.
5. Calvo, F. J., and J. de Oña. Opening of the European Railroad Network: A Lost Opportunity for European Unification. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1916, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 8–19.
6. Sánchez-Borràs, M., C. Nash, P. Abrantes, and A. López-Pita. Rail Access Charges and the Competitiveness of High-Speed Trains. *Transport Policy*, Vol. 17, 2010, pp. 102–109.
7. Crozet, Y. Infrastructure Charging Within the French Railways Sector: A New Challenge. Presented at 11th World Conference on Transportation Research, Berkeley, Calif., 2007.
8. Ramsey, F. A Contribution to the Theory of Taxation. *Economic Journal*, Vol. 37, No. 1, 1927.
9. Boiteux, M. Sur la gestion des monopoles publics astreints à l'équilibre budgétaire. *Econometrica*, Vol. 24, 1956, pp. 22–40. (On the Management of Public Monopolies Subject to Budgetary Constraints. *Journal of Economic Theory*, Vol. 3, pp. 219–240.)
10. González-Savignat, M. Competition in Air Transport: The Case of the High-Speed Train. *Journal of Transport Economics and Policy*, Vol. 38, Part 1, 2004, pp. 77–108.
11. De Rus, G., and V. Inglada. Análisis coste-beneficio del tren de alta velocidad en España. *Revista de Economía Aplicada*, Vol. 1, No. 3, 1993, pp. 27–48.
12. Administrador de Infraestructuras Ferroviarias. *Declaración sobre la red 2010*. 2010. [www.adif.es](http://www.adif.es).
13. European Commission and European Investment Bank. *RAILPAG: Railway Project Appraisal Guidelines*. 2005. [www.eib.org/attachments/pj/railpag\\_en.pdf](http://www.eib.org/attachments/pj/railpag_en.pdf).
14. ECMT. *Les Redevances d'usage des Infrastructures Ferroviaires*. Rapport de la cent septième table ronde d'économie des transports tenue à Paris les 26 et 27 mars 1998. Centre de Recherches Economiques, Paris, 1998.

---

*The Intercity Passenger Rail Committee peer-reviewed this paper.*

# Hierarchical Network Model of Safe High-Speed Rail Operation

Yanhui Wang, Xuemei Xiao, Limin Jia, and Chenchen Zhang

This paper presents a preliminary hierarchical network model based on an analysis of the interaction of relationships between operating safety factors of high-speed rail (HSR). The model is based on the characteristics and factors specific to HSR operation and aims to provide a new theory and method for assessing the safety of HSR operation with the use of the relationship analysis methods of a hierarchical network model. A comprehensive theory of safety assessment in HSR operations is necessary for HSR designers and operators to consider rail safety proactively rather than passively. The development and the application of such a broad-ranging safety assessment theory for HSR operation could help to guarantee efficient HSR operation by protecting the safety of passengers, reducing losses, maintaining social stability, and improving economic efficiency.

There are many advantages to developing high-speed rail (HSR). It has quickly become the first choice for intercity transportation because it not only increases capacity but also provides the safest, most effective, and most environmentally friendly transport means. At present, the world has around 10,700 km of operational HSR, including about 2,000 km in Japan and about 1,900 km in France. Although China has become an HSR leader in the world after several years of rapid development, it must continue to develop its HSR network to raise capacity for its passenger transport needs. China plans on building 42 high-speed passenger rail lines with a total length of 13,000 km by 2012 and 16,000 km by 2020. These lines are named the “four length lines and four breadth lines” of the Chinese HSR network (Figure 1).

The sudden growth of HSR in China has raised issues not merely of passenger demand for transportation services and increased speed (decreased travel times) but also of the integration of transport organization and management and, most importantly, safety controls.

While clearly safety should be considered in the planning, design, and construction of any HSR system, it should also be integral to operation management. Yet studies of individual factors' effects on the system are insufficient to provide a comprehensive understanding and means of modifying the complexities of HSR safety. Emergency mechanisms should be studied from a systemwide perspective to find potentially unsafe factors, to define a fault linkage model, and

to provide a theoretical basis for the safe operation management and maintenance of infrastructure and mobile devices. Although some studies have analyzed the safety of HSR operation, they generally favor technical considerations over theory, and qualitative analysis has predominated over quantitative calculations; moreover, the studies that focus on safety have tended to be local and microlevel rather than global and macrolevel (1). In short, the unsafe practices of HSR operation have not been comprehensively studied at a systemwide level, and there is little research on the interaction mechanism of safety factors, which is necessary for a thorough safety assessment of HSR operation.

A study of the safety mechanisms of HSR operation and their interactions and the development of a comprehensive theory of safety assessment in HSR operations are necessary for HSR designers and operators to consider rail safety proactively rather than passively. The development and application of such a broad-ranging safety assessment theory for HSR operation could help to guarantee efficient HSR operation by protecting the safety of passengers, reducing losses, maintaining social stability, and improving economic efficiency.

This paper responds to this need by offering a preliminary hierarchical network model that can describe the interaction relationships between safety factors of the HSR operation. The model, which is based on an analysis of characteristics and factors specific to HSR operation, aims to provide a new theory and method for the safety assessment of HSR operation by using the relationship analysis methods of a hierarchical network model.

## CHARACTERISTICS OF HSR OPERATION

Unlike the conventional railway, HSR is a large, dynamic, complex system with a high degree of technical complexity, complex equipment, high integration, and a high coupled degree in each subsystem. There are clear differences in the operation rules, properties, and interaction with the environment between conventional and HSR. Unlike the conventional railway, the HSR system has characteristics of complex systems, including nonlinearity, uncertainty, high-coupled conditions, high dimension, and chaos.

### High Speed

The high speeds of HSR trains result in large kinetic energy, and its impact force on the rolling stock and a variety of forces that the trains incur during operation are greatly improved (2). However, the train's inertia is also large, and a considerable distance is required to successfully stop or slow the train with the emergency brake when obstacles are encountered on the track. In addition, the train is easy to derail.

State Key Laboratory of Rail Traffic Control and Safety, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China. Corresponding author: X. Xiao, xiaoxuemei1983@yahoo.com.cn.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 49–56.  
DOI: 10.3141/2261-06

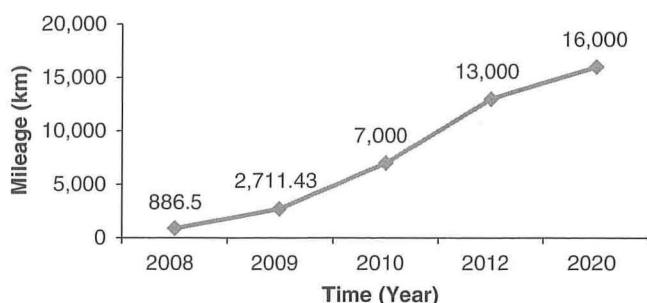


FIGURE 1 China's HSR development planning.

### Complex Technology and High Integration

HSR is a complex integrated and coordinated system composed of bridges, tunnels, rolling stocks, high-speed trains, signaling, micro-electronics, computers, communication, control, new materials, and various technologies. It is the largest and most complex modern system after the aerospace industry.

### High Coupled Degree

The coupled degree of subsystems and the internal components within each subsystem is high; the interaction relationship for HSR is nonlinear and more complex than for conventional rail.

### Organization

Conventional rail has traditionally relied on human and organizational systems to guarantee safety. Because of the high speed and large density of HSR, many factors are necessary to maintain its operation. HSR safety requires holistic organization management, safety information sharing, and overall decision-making support.

## ANALYSIS OF FACTORS IN SAFE HSR OPERATION

HSR needs to operate in a high-safety environment because a small breach could lead to an accident when the train travels at full speed. From the view of safety system engineering, human, facility, environmental, and management factors affect safe HSR operation. HSR operation accidents occur largely as a result of unsafe human behavior, the unsafe state of the facilities required by HSR operations, unsafe environmental conditions, and poor management (Figure 2). These four factors share the significant characteristics of time and space distribution.

### Human Factors

Safe HSR operations demand that both technology and management personnel should adapt to new situations and master new technology as needed. Additionally, they should have a high level of technical capabilities, the ability to respond quickly, a serious working attitude, and excellent performance.

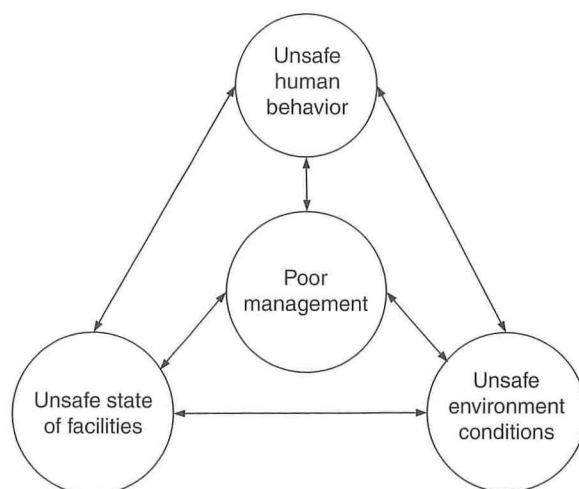


FIGURE 2 Factors affecting safe HSR operation.

### Facilities Factors

Good performance of facilities, such as vehicles, steel rail, signaling, and catenary–pantograph, is the basis for HSR and is indispensable in guaranteeing safe operation. Newer facilities tend to have higher reliability and security, but with longer use their safety performance decreases as a result of physical and chemical factors (e.g., wear, corrosion, fatigue, and aging). Accident analysis indicates that the following factors affect safe HSR operation.

#### Steel Rail

The smoothness of the steel rail surface is directly related to safe operation. Even a small irregularity can cause strong vibrations in a high-speed train that lower its comfort, stability, and safety. As a result of the high density of HSR trains, steel rail is prone to wear and fracture under repeated train loads. Additionally, increasing rail temperatures increase the longitudinal stress of the welded rail. High levels of steel rail expansion occur when the rail temperature increases past a certain threshold value, and this expansion can affect safe operation.

#### Roadbed

Roadbed is the basis of steel rail. The roadbed of HSR must have sufficient strength, stability, and durability to resist the impact of natural factors in order to guarantee safe and smooth running. The ballasted track is sensitive to roadbed settlement, particularly to differential settlement.

#### High-Speed Train

Various factors affect high-speed passenger trains, including the following:

1. Speed. The higher the train speed is, the bigger its vibration acceleration is. The direct consequence of increased vibration

acceleration is looseness of rail joint accessories and a decrease in the stability of track geometry. With increased train speed, in addition to the increase in vertical force  $P$  of the wheel track, the lateral force  $Q$  increases correspondingly; as a result, the derailment coefficient  $K = Q/P$  increases, which can lead to a decrease in train safety.

2. Traffic volume. Traffic volume is the sum of train traction tonnage through a given line within a specified time. It usually equals the product of the train axle load and traffic density. The higher the traffic density is, the larger the traffic volume is, the bigger the number of wheels in unit time is, and the faster the rail wears. This would affect the smoothness of the steel rail. Both factors give rise to an increase of track maintenance costs (3).

3. Bogie. Bogie is a support of the rail vehicle body that ensures a comfortable ride by absorbing vibration and minimizing centrifugal forces when the train runs on curves at high speed. The partial load of the bogie affects the safety of the vehicle. As the partial load increases, the vehicle's safety level decreases.

4. Train traction equipment. The leakage, arc, electric spark, static discharge, and electrical fire of the electrical installations in vehicles can cause accidents such as fires (4, 5).

### *Catenary–Pantograph*

The locomotive power supply of high-speed trains is provided through the interaction of the catenary and pantograph. Fast, safe, and efficient operation of vehicles depends on the performance of the catenary–pantograph system. As the speed of an HSR train increases, the dynamic changes of the catenary increase, and there will be off-line faults between catenary and pantograph. The catenary–pantograph system can also be affected by the outside environment, such as winds (6).

### *Signaling*

Traffic signaling systems (command and control equipment) play an important role in ensuring safe operation, especially in monitoring train speed, maintaining spacing distance, and protecting against overspeed in real time.

### *Communication*

Communications systems and equipment exchange and transmit information between high-speed trains and wayside or central offices in real time and form the key feature in realizing automation and centralized command and control of HSR operation.

### *High Bridges and Tunnels*

The strict design parameters of HSR require a large curve radius and a small slope. Because the high-speed train should operate in a closed environment, many high bridges and tunnels are required. The dynamic performance of these bridges and tunnels affects safe HSR operation. For example, the principal dynamic concerns of a bridge that are related to dynamic load factor are structural design, ballast stability, carriage acceleration, and radiated noise (7).

## **Environmental Factors**

Although the high-speed train G2003, G2007 bound from Zhengzhou to Xi'an in China was delayed because of equipment failure caused by bad weather on February 7, 2010, in general weather conditions (with the exception of strong winds) have little influence on HSR. In contrast, natural disasters (such as earthquakes and floods) can seriously affect its safe operation.

### *Strong Wind*

Strong winds can cause vibration and oscillation of transmission lines and the catenary–pantograph system as well as wind vibration in large span bridges. Crosswinds can also threaten high-speed trains as they run over high bridges (8).

### *Hydrologic Risk*

Hydrologic risk includes rainfall, flood, and landslides. Intense rainfall may affect the driver's visibility and thus decrease the train's speed and cause the operation to be delayed. Flooding and landslides may lead to roadbed settlement or the breach of the steel rail, and even directly result in an accident (9).

### *Earthquake*

Earthquakes are devastating natural phenomena that can produce strong vibrations in steel rail, destroy bridges, and directly affect safe high-speed train operation. The earthquake in Niigata Ken Chuetsu, Japan, on October 23, 2004, for example, caused extensive damage to roadbeds as a result of ground failures (10, 11).

### *Invasion*

Although strictly speaking invasion is not an environmental factor, it would almost certainly affect the safety of power supply facilities and thus endanger HSR operation. Any obstacles on the track as a result of invasion could increase the chance of derailment because of the difficulty of stopping a high-speed train with the train emergency brake.

## **Management Factors**

As shown in Figure 2, management is central to safe HSR operation. It is management's responsibility to oversee the inspection and maintenance of all critical equipment and facilities, judge the operational state of all components of the HSR system, warn of possible adverse events, and establish an HSR emergency system; in the future, an additional responsibility will be to realize information sharing of HSR (12, 13).

Figure 3 shows how the four main factors discussed above interact with each other and affect the safe operation of HSR systems. The analysis of the four main factors and their subfactors is preliminary. In fact, each unit comprises a number of components that interact with and influence each other. These additional interactions are not discussed here.

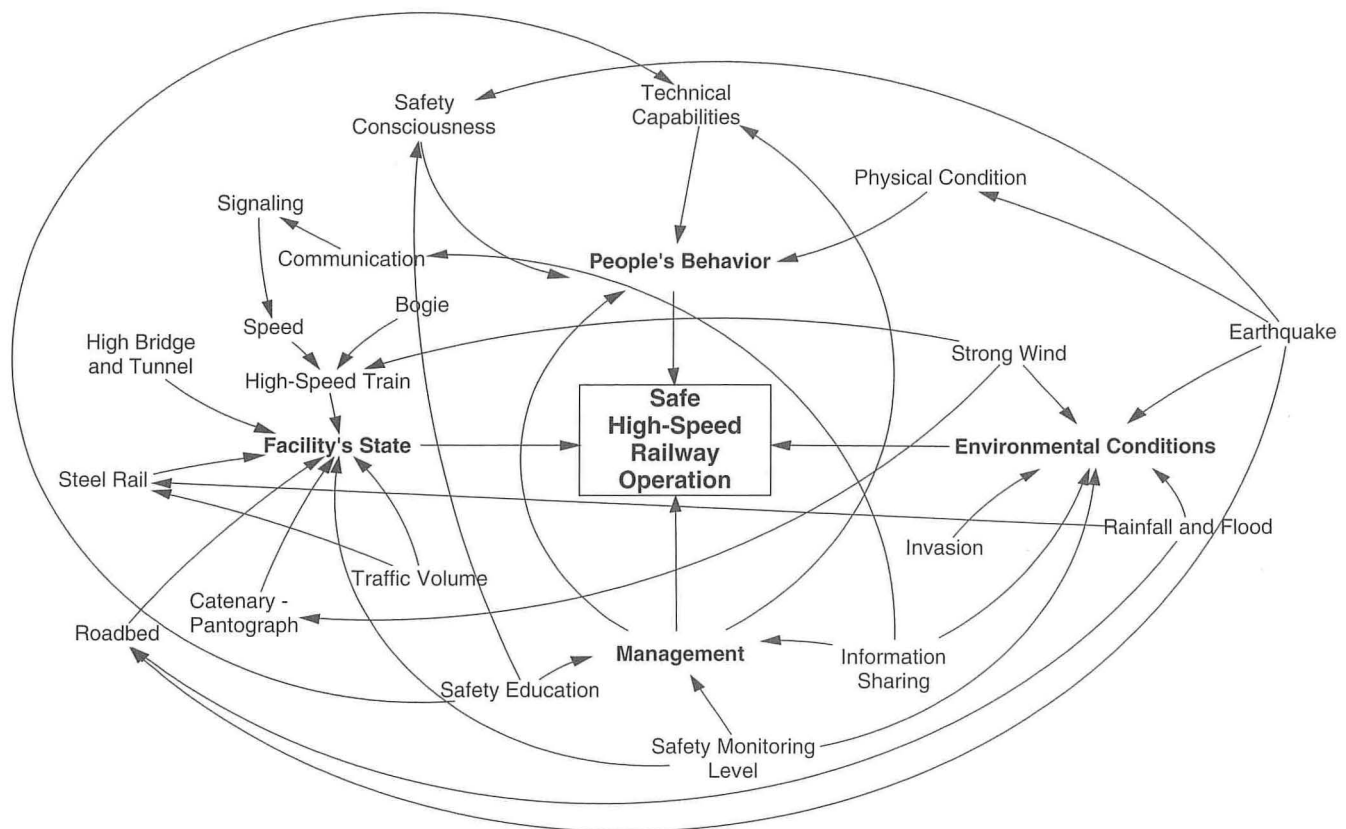


FIGURE 3 Interaction between factors.

## HIERARCHICAL NETWORK MODEL OF SAFE HSR OPERATION

### Complex Network

From the view of systems engineering, the units that affect safe HSR operation are interrelated and coordinated to exchange materials, information, and energy. A breach in any one of the basic factors in any subsystem of an HSR operation is likely to spread, accumulate, magnify, and eventually result in an accident. A close study of the occurrence, interaction mechanisms, and dynamic evolvement of factors that affect safe HSR operation requires an accurate, effective, and quantitative method that allows the intrinsic system rules to be observed and controlled.

The traditional qualitative or quantitative evaluation methods, such as failure mode and effects analysis and fault tree analysis, cannot satisfy the safety analysis requirements of a complicated HSR system. Failure mode and effects analysis calls for some subjective judgments, and there is a lack of systematic and thorough description of the system risk. Fault tree analysis is problematic because the fault tree of a complex system such as HSR is very large, and the amount of calculation is very great.

A complex network, as an important part of a complex system, can describe the topology and behavior of a complex system. Based on statistical physics, nonlinear dynamics system theory, and complexity science, a complex network can model and analyze complex systems. Complex networks have been widely applied in social economics, computer networks, engineering, neuroscience, medicine, and transportation (14).

### Hierarchical Network Model

Graph theory is the natural framework for the discrete mathematics of complex networks; formally, a complex network can be modeled as graphs. According to the structure and function of the HSR system and based on complex network theory and methods, the factors that affect the safe operation of HSR and their relationships were mapped as network topology. This process entails translating the interaction between two dynamic units, which depend on time, space, and many other details, into a simple binary number. A safety hierarchical network model of HSR operation that can describe the interaction relationships between safety factors was developed in the present study (Figure 4) (15).

In the basic layer, according to graph theory, the factors that affect safe HSR operation are regarded as the nodes of the network, and their interaction relationships are regarded as the adjacent arc. The formalized definitions of the network model are listed below.

#### Definition 1. Safety Factor

At the bottom of the model, each node represents a factor that influences safe HSR operation, that is,  $s_{ij}$  denotes the  $j$ th safety factor in the  $i$ th subsystem. Each safety factor is an independent safety object, with different dynamic behavior, uncertainty, and randomness properties;  $x_k(s_{ij})$  denotes the state of  $s_{ij}$  at different times. Through the analysis, the factors that affect safe HSR operation were classified into two types: continuous variables, such as velocity, accelerated

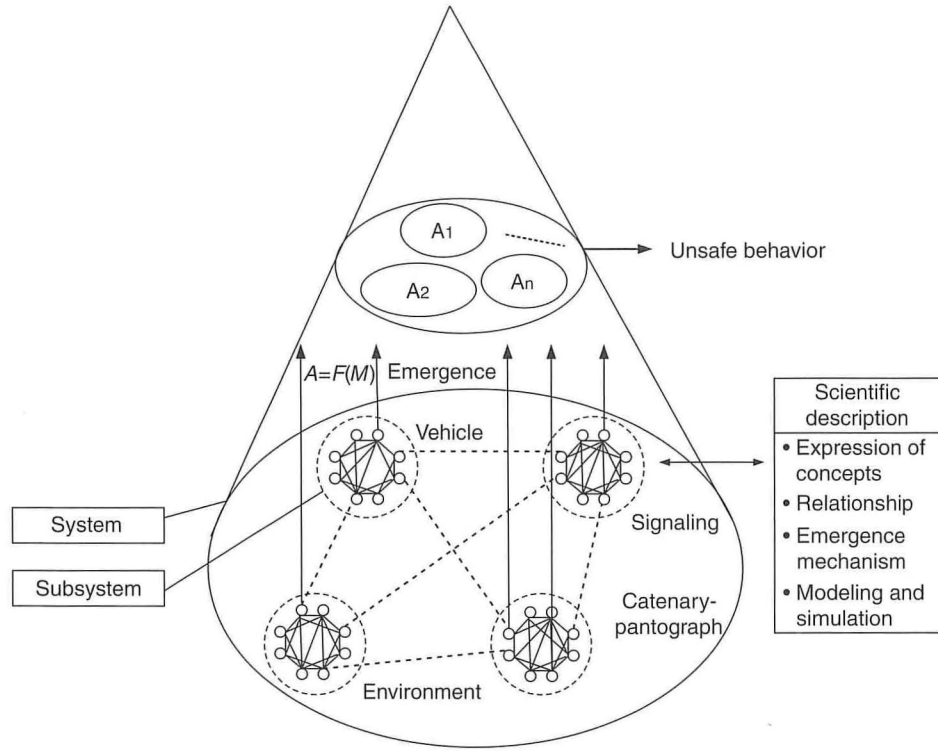


FIGURE 4 Hierarchical network model of safe HSR operation.

velocity, and temperature; and discrete variables, such as signaling, interlock, and alarm.

### Definition 2. Safety Unit

The safety unit is composed of safety factors interrelated in a certain way (i.e., with a time correlation and space correlation):

$$S_i = \{s_{ij}\} \quad (1)$$

The safety unit includes human factors set  $S_1$ , vehicle factors set  $S_2$ , rail factors set  $S_3$ , catenary–pantograph factors set  $S_4$ , signaling factors set  $S_5$ , communication factors set  $S_6$ , environmental factors set  $S_7$ , and management factors set  $S_8$ .

### Definition 3. Safety Factors Set

The safety factors set is composed of all safety factors that influence the safe operation of HSR:

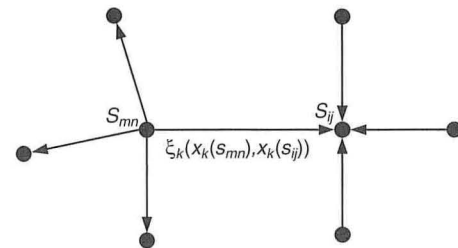
$$S = S_1 \cup S_2 \cup \dots \cup S_i \quad (2)$$

The safety factors set of safe HSR operation is  $S = S_1 \cup S_2 \cup S_3 \cup S_4 \cup S_5 \cup S_6 \cup S_7 \cup S_8 = \{s_{mn}\}$ , where  $u = 1, 2, \dots, 8$  and  $v = 1, 2, \dots, j$ . Although the safety factors are independent, coupled interactions exist between them. As shown in Figure 5, the direction of the edge stands for the cause-and-effect relationship of the coupled interaction. For example, a change in node  $s_{mn}$  (denoting the  $m$ th safety factor in the  $n$ th subsystem) may lead to a change in  $s_{ij}$ , that is, the basic layer is a weighted network. In a weighted network, the coupled strength

between the nodes is the key factor that restricts its dynamic behaviors. In this paper, the coupled strength between the nodes is calculated by gray correlation analysis. Because the physics meaning or measuring unit of each safety factor differs, their dimensions are different, and sometimes their order of magnitude is bigger. Therefore, it is necessary to process the original data so it is dimensionless; this process is not detailed here.

Suppose the safety factors that have coupled interaction with  $s_{ij}$  are as shown in Figure 5, and suppose the sequences of each safety factor at different times after the quantization process are

$$\begin{aligned} X(s_{ij}) &= \{x_1(s_{ij}), x_2(s_{ij}), \dots, x_k(s_{ij})\} \\ &\dots \\ X(s_{mn}) &= \{x_1(s_{mn}), x_2(s_{mn}), \dots, x_k(s_{mn})\} \\ &\dots \\ X(s_{m+1,n+1}) &= \{x_1(s_{m+1,n+1}), x_2(s_{m+1,n+1}), \dots, x_k(s_{m+1,n+1})\} \end{aligned} \quad (3)$$

FIGURE 5 Safety factors that have coupled interaction with  $s_{ij}$ .

#### Definition 4. Correlative Coefficient

The correlative coefficient is the measurement of the influence of  $s_{mn}$  on  $s_{ij}$  at time  $k$ :

$$\xi_k(x_k(s_{mn}), x_k(s_{ij})) = \frac{\min_{s_{mn} \in N_{ij}} \min_k |x_k(s_{ij}) - x_k(s_{mn})| + \rho \max_{s_{mn} \in N_{ij}} \max_k |x_k(s_{ij}) - x_k(s_{mn})|}{|x_k(s_{ij}) - x_k(s_{mn})| + \rho \max_{s_{mn} \in N_{ij}} \max_k |x_k(s_{ij}) - x_k(s_{mn})|} \quad (4)$$

where

$$\begin{aligned} \xi_k(x_k(s_{ij}), x_k(s_{mn})) &= \text{correlative coefficient between } s_{ij} \text{ and } s_{mn} \text{ at time } k, \\ x_k(s_{ij}) &= \text{state of } s_{ij} \text{ at different times, and} \\ \rho &= \text{distinguishability to reduce the influence on the calculation of extreme value, } 0 < \rho < 1. \end{aligned}$$

#### Definition 5. Coupled Strength

The correlative coefficient only stands for the coupled interaction between two safety factors at time  $k$ . In order to know the coupled strength overall, the coupled strength between the two safety factors is the average of the correlative coefficients at time  $\{1, 3, \dots, t\}$ .

$$r(X(s_{mn}), X(s_{ij})) = \frac{1}{t} \sum_{k=1}^t \xi_k(x_k(s_{ij}), x_k(s_{mn})) \quad (5)$$

where  $r(X(s_{mn}), X(s_{ij}))$  is the coupled strength between  $s_{ij}$  and  $s_{mn}$ , and  $\xi_k(x_k(s_{ij}), x_k(s_{mn}))$  is the correlative coefficient between  $s_{ij}$  and  $s_{mn}$  at time  $k$ .

The formalized description of the basic layer of the hierarchical network model can be expressed by  $G=(S, R)$ ; the adjacency matrix is used to represent the weighted network.

$$R = \begin{pmatrix} r(X(s_{11}), X(s_{12})) \dots r(X(s_{11}), X(s_{1n})) \\ \vdots & \dots & \ddots & \vdots \\ r(X(s_{ij}), X(s_{12})) \dots r(X(s_{ij}), X(s_{mn})) \end{pmatrix}$$

$$\text{when } \begin{cases} r(X(s_{ij}), X(s_{mn})) \neq 0, \\ \text{there is coupled interaction between } s_{mn} \text{ and } s_{ij} \\ r(X(s_{ij}), X(s_{mn})) = 0, \\ \text{there is no coupled interaction between } s_{mn} \text{ and } s_{ij} \end{cases} \quad (6)$$

where  $R$  is the adjacency matrix, and  $r(X(s_{mn}), X(s_{ij}))$  is the coupled strength between  $s_{ij}$  and  $s_{mn}$ .

In network topology, there are usually a few nodes with many connected edges and many nodes with few connected edges, assembling characteristics for certain nodes in the network.

#### Definition 6. Node Strength

Node strength is the indicator that quantitatively describes the importance of each factor in the network. In the basic layer of the hierarchical network model, the node strength of  $s_{mn}$  is the sum of the coupled strengths of its direct connections to other nodes (Figure 5).

$$k(s_{mn}) = \sum_{s_{ij} \in N(s_{mn})} r(X(s_{mn}), X(s_{ij})) \quad (7)$$

where

$$\begin{aligned} k(s_{mn}) &= \text{node strength of } s_{mn}, \\ r(X(s_{mn}), X(s_{ij})) &= \text{coupled strength between } s_{ij} \text{ and } s_{mn}, \text{ and} \\ N(s_{mn}) &= \text{node set coupled with } s_{mn}. \end{aligned}$$

The higher the node strength is, the greater the effect on the stability of the network when the node breaks down. It is a key node in the model.

According to accident chain theory, an accident is not an isolated incident but rather a series of cause-and-effect events. Any safety factor's unsafe behavior or failure may spread and eventually lead to an accident. In other words, an HSR operation accident is a series of fault chains.

#### Definition 7. Propagation Paths

Different propagation paths correspond to different HSR operation accidents. In this paper propagation paths in the basic layer are mapped to show unsafe behaviors.

$$P = \{s_{ij} \rightarrow s_{i+1,j} \rightarrow \dots \rightarrow s_{mn}\} \rightarrow A_i \quad p \rightarrow A \quad (8)$$

where  $P$  is the propagation path, and  $A_i$  is the safety state. Propagation paths depend on the propagation rate between two points. Because the coupled strength between the factors changes over time, the path changes also. There are three types of fault chain; the derailment of an HSR train is used for an example of each type:

1. Single-series accident chain (Figure 6),
2. Single parallel accident chain (Figure 7), and
3. Composite structure accident chain (Figure 8).

#### Definition 8. Propagation Rate

If coupled interactions exist between  $s_{ij}$  and  $s_{mn}$  and if  $s_{mn}$  changes, then the probability that  $s_{ij}$  is affected is expressed by

$$\lambda(X(s_{mn}), X(s_{ij})) = \frac{r(X(s_{mn}), X(s_{ij}))}{k(s_{ij})} = \frac{r(X(s_{mn}), X(s_{ij}))}{\sum_{s_{mn} \in N(s_{ij})} r(X(s_{mn}), X(s_{ij}))} \quad (9)$$

where  $\lambda(X(s_{mn}), X(s_{ij}))$  is the propagation rate.



FIGURE 6 Single-series accident chain.



## CONCLUSION

By analyzing the factors that affect safe HSR operation, the authors constructed a preliminary hierarchical network model of safe HSR operation that can describe the interaction relationships between safety factors. The knowledge representation and relationship analysis methods of the hierarchical network model were studied in detail. By using complex network and safety entropy theory, both local and global safety assessments of HSR operation can be realized. This paper sheds light on how an HSR operation accident is the emergence of the coupled interactions between safety factors. In the future, these emergence and evolution mechanisms are to be explored in depth to provide a new theory and a new method for safety evaluation of HSR operation.

## ACKNOWLEDGMENTS

This work was supported by grants from the Theory Study on Rail Traffic System Safety Analysis Based on Complex Systems and from funding for discipline and graduate education construction projects from the Beijing Municipal Commission of Education.

## REFERENCES

1. Junwei, G., C. Guoqiang, Q. Yong, and J. Limin. Safety Analysis of Rail Traffic Dynamic System Based on Mixed Variable Space (in Chinese). *Urban Mass Transit*, No. 3, 2009, pp. 22–24.
2. Jing, Z., and W. Ping-Bo. Stability of High-Speed Train (in Chinese). *Journal of Traffic and Transportation Engineering*, Vol. 5, No. 2, 2005, pp. 1–4.
3. Teixeira, P. F., A. López-Pita, C. Casas, A. Bachiller, and R. Robusté. Improvements in High-Speed Ballasted Track Design: Benefits of Bituminous Subballast Layers. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1943, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 43–49.
4. Paul, M. *Fire Risk Assessment Study for a High-Speed Train*. PMSC Limited Paper on Generic Approach to Fire QRA for High Speed Trains. PM Safety Consultants Limited, Coventry, United Kingdom.
5. Hagiwara, Y., S. Ishikawa, and M. Furuya. Innovative Lightweight Technologies Using Power Electronics on Shinkansen High-Speed Electric Multiple Units. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1995, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 43–51.
6. Yuan, Y., W. Junyong, and Z. Xiaoyu. RAMS Evaluation of a Bulk Power System to a Traction Power Supply System of High-Speed Rail. *Automation of Electric Power Systems*, No. 20, 2007, pp. 98–102.
7. Willford, M., K. Hicyilmaz, and K. Pratt. Dynamic Considerations in the Design of High-Speed Railway Bridges. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1863, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 28–36.
8. Zunsong, R., X. Yugong, W. Lu-lei, and Q. Ying-Zheng. Study on the Running Safety of High-Speed Trains Under Strong Cross Winds (in Chinese). *Journal of the China Railway Society*, Vol. 28, No. 6, 2006, pp. 46–50.
9. Veneziano, Daniele. Hydrologic Risk for High Speed Rail System. 2008. [http://www.mitportugal.org/index.php?option=com\\_docman&task=cat\\_view&gid=115&limit=100&limitstart=0&order=hits&dir=DESC&Itemid=1](http://www.mitportugal.org/index.php?option=com_docman&task=cat_view&gid=115&limit=100&limitstart=0&order=hits&dir=DESC&Itemid=1).
10. Xuesong, J., G. Jun, X. Xin-Biao, W. Zefeng, and Z. H. Zhongrong. Key Scientific Problems in the Study on Running Safety of High-Speed Trains (in Chinese). *Engineering Mechanics*, No. S2, 2009, pp. 8–22.
11. Preliminary Observations on the Niigata Ken Chuetsu, Japan, Earthquake of October 23, 2004. EERI Special Earthquake Report. Earthquake Engineering Research Institute, Oakland, Calif., 2005.
12. Marcolongo, V., W. Ebersöhn, and Michael J. Trosino. Information Requirements for Railroad Enterprise Asset Management: Practical Approach on Amtrak's High-Speed Northeast Corridor Railway Line. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1863, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 74–80.
13. Park, H., B. L. Smith, R. Venkatanarayana, and H. Song. Exploration of Operational Concepts of Interoperability for Robust Interregional Transportation System Operations. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2178, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 138–146.
14. Xingtang, L., H. Bingcheng, and L. Li. *The Theory, Method and Technique for Complex System Modeling* (in Chinese). Science Press, Beijing, 2008.
15. Shiyong, L. *Nonlinear and Complexity Science* (in Chinese). Harbin Institute of Technology Press, Harbin, China, 2006.
16. Leveson, Nancy G. *Engineering a Safer World: System Safety for the 21st Century*. Draft book. Massachusetts Institute of Technology Press, Cambridge, 2009.
17. Tianfu, C., and Z. H. Jinglin. Discussion on Operational Mechanism of Safety Systems: Degree of Safety and Entropy of Safety. *China Safety Science Journal*, Vol. 16, No. 3, 2006, pp. 4–7.
18. Jinglin, Z. H., and C. Tianfu. Discussion on Operational Mechanism of Safety Systems: Essential Character and Structure of Safety Systems. *China Safety Science Journal*, Vol. 16, No. 5, 2006, pp. 16–21.

---

*The Passenger Rail Equipment and Systems Integration Committee peer-reviewed this paper.*

# Simplified Approach for Assessing Initial Fire Development and Spread in Passenger Rail Vehicles

B. J. Meacham, N. A. Dembsey, M. Johann, J. Tubbs, and K. Schebel

Whether a fire in a passenger rail vehicle remains small or grows to encompass the vehicle and how large the fire may become largely depend on the initiation fire, vehicle interior lining materials and contents, vehicle configuration, and ventilation. Assuming fixed configuration and ventilation, the initiation fire, interior materials, and contents are critical variables. Evaluation of the fire hazards represented by different interior lining materials when exposed to various initiation fire scenarios can be costly and time-consuming with full-scale fire tests. However, small-scale fire tests are less expensive, and numerous tests can be performed in a short time. Small-scale test data coupled with initiation fire data and computational modeling can be used to cost-effectively assess a wide range of scenarios and material combinations for existing and proposed vehicle designs. This paper proposes a simplified approach for characterizing initiation fires and predicting the potential for flame spread on the basis of data that can be obtained from small-scale fire tests. Through the application of a simple screening tool, it is possible to quickly assess a material's propensity to spread flames over a range of initiation fires. This approach can be used to support threat, risk, and vulnerability assessments of existing rail stock and to screen materials for new vehicle design. In conjunction with additional assessment of initial flame spread characteristics, the screening tool can be used with computational fluid dynamics fire-modeling tools to predict the overall development and spread hazard of the fire for specific vehicle configurations and resultant infrastructure impacts.

Fire in a passenger rail vehicle is influenced by the size of the initiation fire; the type, amount, and characteristics of fuel available to burn; the size and location of ventilation openings; and the compartment (vehicle) configuration. For any given combination of these parameters, the size and location of the initiating fire can significantly affect the fire size. Factors such as the effects of adjacent materials on fire spread and growth, interaction of the fire with the geometric elements of the car, and the potential for flame spread away from the fire origin must be considered. Developing a better understanding of the relationship between initiation fire and interior

lining materials and the influence of these on the maximum fire size that may be expected is critical to fire hazard analysis for new and existing vehicles. Such information is essential for estimating how large a vehicle fire might grow, which will better inform threat, risk, and vulnerability assessments (TRVAs) of vehicles and critical rail infrastructure.

Although existing regulations require testing of material flammability and flame spread (1, 2), the tests used are not conducive to informing TRVA as there is no connection between initiation fire and material response. As an initial step toward a more comprehensive fire hazard assessment methodology and to enhance current TRVA approaches, research into the characterization of initiation fires and prediction of interior flame spread given the initiation fires was conducted under a grant from the International Programs Division, Science and Technology Directorate, Department of Homeland Security. This effort focused on the relationship between initiation fire size and interior lining materials, with an emphasis on understanding initiation fires, the range of interior materials that are commonly used, and the size of initiation fire that can lead to self-propagating fires given the combustible interior materials used. This research effort also resulted in the development of a simplified flame spread model and an approach to incorporate flame spread characteristics into a computational fluid dynamics (CFD) model of the vehicle. The process of collecting material property data, screening for flame spread potential, and incorporation of outcomes into CFD analysis is illustrated in Figure 1 (3). For existing stock, this process provides a means to quickly assess what size initiation fire could lead to a fully developed fire in a vehicle, and for new vehicles, it provides an enhanced tool for assessing fire performance of materials.

This paper focuses on initial threat assessment components, including characterization of the initiation fire, the use of small-scale test data to assess the propensity for interior materials to self-propagate a fire given a range of initiation fires, and how this information can be used in TRVA. The simplified flame spread model and the incorporation of the flame spread model output into CFD analysis for use in infrastructure vulnerability assessment are described in detail elsewhere (3; Meacham et al., unpublished work).

## CHARACTERIZATION OF INITIATION FIRES

TRVA for rail vehicles includes assessing the first materials burning (initiation fire), the secondary materials and their propensity for spreading the fire, and the impacts of the resulting fire development and spread (1, 2, 4). A first step in the process is understanding and characterizing the initiation fires and scenarios of concern. As part

---

B. J. Meacham and N. A. Dembsey, Department of Fire Protection Engineering, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609. M. Johann and J. Tubbs, Arup USA, 955 Massachusetts Avenue, 4th Floor, Cambridge, MA 02139. K. Schebel, Arup USA, 155 Avenue of the Americas, New York, NY 10013. Corresponding author: B. J. Meacham, bmeacham@wpi.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 57–63.  
DOI: 10.3141/2261-07

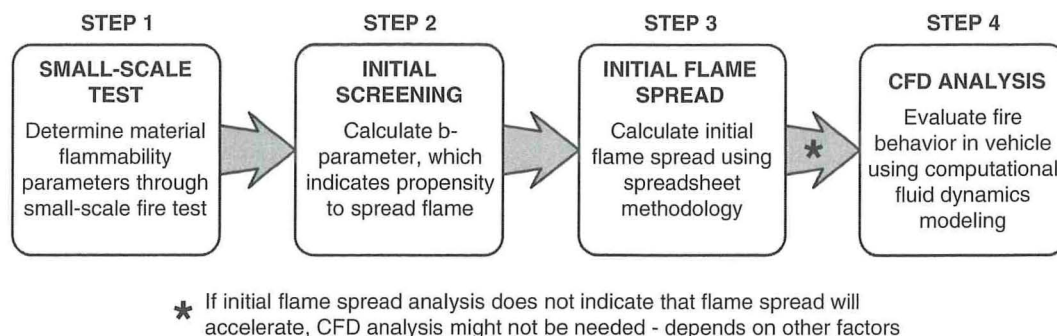


FIGURE 1 Process for assessing potential for flame spread in vehicles.

of this effort, data were collected from published reports and the media to inform the range of initiation fire scenarios, including common ignition sources, initial materials, initial fire sizes, and vehicle materials involved (3, 5, 6). Review of these data suggests that a majority of the fires were accidental and originated outside of train compartments. Nonetheless, from 2003 to 2008, some 272 fire-related injuries were reported (5). Even though most accidental fires initiate outside of vehicles, inside initiation fires have resulted in severe damage, and although less frequent, intentional interior fires have resulted in some of the most significant losses of life and physical assets. The Daegu, South Korea, subway fire of February 18, 2003, for example, resulted in the deaths of 198 people, injured at least 147, and caused significant infrastructure impacts. The Daegu arsonist was reported to have used two milk cartons filled with flammable liquids for the initiation fire. The fire spread quickly, and six cars were fully engulfed within 2 min.

In interior fires, the type, quantity, arrangement, and location of the first materials ignited are important factors in determining if the initiation fire will lead to the involvement of interior lining materials. In particular, these factors are important with respect to radiant ignition of nearby materials, which is a function of the energy output of the initiation fire, flame height and geometry, the distance to the

secondary fuel source, and the ignition characteristics of the secondary fuel. Figure 2 depicts typical flame heights, energy (heat) release rates, and heat flux levels resulting from a range of items burning in the open (unconfined).

In Figure 2, heat flux levels are measured approximately 0.3 m (1.0 ft) horizontally from the fire. Data for the trash bags, sleeping berth, and cushioned seats come from fire tests associated with passenger rail materials; the office workstation is included for comparison (6). It is possible to obtain fires of 2,000 kW and above for liquid pool fires, depending on the fuel, pool diameter, and depth. Thus, these fires can be considered representative of initiation fires that could be realized in passenger compartments. For a particular TRVA, data could be collected and characterized with actual fire history within a rail system and knowledge about materials transported on the system.

With an understanding of the size of initiation fires, it is important to then characterize them in such a way as to assess the likelihood of fire spread. In general, the issues of concern are intensity (temperature and radiant heat flux) and duration of fire in relation to the material that may next become ignited (e.g., a match can easily ignite a thin piece of paper, but will not be able to ignite a thick block of wood). In order to quantify whether an initiating fire is likely to

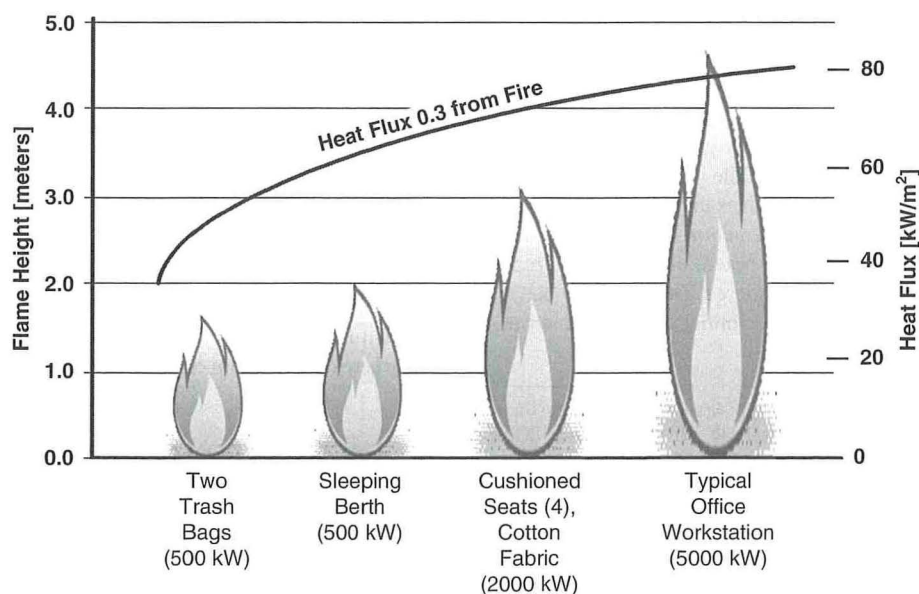


FIGURE 2 Relative fire sizes and heat fluxes 0.3 m horizontally from fire.

spread to adjacent lining materials, the thermal exposure resulting from the initial fire needs to be reviewed. To evaluate whether a material will ignite under a given fire exposure, the material's critical heat flux (CHF) must be considered. CHF is defined as the minimum heat flux at or below which a material does not generate vapors to support combustion. The parameters used to calculate CHF are external heat flux and time to ignition, both of which can be controlled or derived with the cone calorimeter.

The cone calorimeter, a standardized testing apparatus (7–9) that uses oxygen consumption calorimetry to measure the heat release rate of a small-scale material specimen (10, 11), is a widely used bench-scale fire testing apparatus for evaluating a material's flammability. Cone calorimeter tests involve exposing a specimen measuring  $0.1 \times 0.1 \text{ m}$  ( $\sim 0.01 \text{ m}^2$ ) to a constant external radiant flux by passing an electrical current through the cone-shaped heating element, where the radiant flux incident to the sample surface can be varied between 0 and  $100 \text{ kW/m}^2$ . This process allows the evaluation of fire performance and burning characteristics, such as heat release rate per unit area (HRRPUA) and CHF, under different heat flux conditions (7). Given the CHF for a material, thermal exposures calculated from cone calorimeter data can be compared with expected thermal exposures for scenarios of concern. This comparison provides an approach to assess whether a given initiation fire is likely to ignite secondary materials.

Briefly, the thermal exposure within the cone calorimeter can be calculated with the following formula:

$$(\dot{q}'' - \text{CHF}) * t_{\text{ig}} \quad (1)$$

where  $\dot{q}''$  is the HRRPUA in kilowatts per meter squared, and  $t_{\text{ig}}$  is the ignition temperature ( $^{\circ}\text{C}$ ). With a fiber-reinforced plastic (FRP) gel coat (paint) as an example, the thermal exposures at multiple external heat fluxes can be determined by cone calorimetry as illustrated in the table below:

External Heat Flux ( $\text{kW/m}^2$ )	Ignition Time (s)	Thermal Exposure ( $\text{kJ/m}^2$ )
25	447	2,235
50	246	7,380

Known factors of external heat flux, thermal exposure, and time to ignition for a material can be compared with the characteristics of the initiation fire to assess whether a competent ignition source exists (i.e., an initiation fire that burns sufficiently long and with enough energy to ignite the secondary material). Heat flux exposures to lining materials will depend on the item burning and its proximity to the lining materials. Scenarios of interest might include trash bags, suitcases, and vehicle seats (6) as well as flammable liquids. Thermal exposures for representative initiation scenarios are shown in Table 1.

**TABLE 1 Thermal Exposures from Fire Scenarios**

Initiation Fire Scenario	Thermal Exposure ( $\text{kJ/m}^2$ )
Trash bags	4,029
Hard suitcase	5,625
Soft suitcase	NA
Sleeping berth	9,835
Double seat	8,173

NOTE: NA = not applicable.

Using thermal exposure as a comparative tool, one can see from Table 1 that every scenario, except that of the soft suitcase, resulted in a thermal exposure greater than that indicated by the cone calorimeter data collected at a heat flux of  $25 \text{ kW/m}^2$ . In the case of the soft suitcase, thermal exposure values in excess of CHF were not observed, so ignition of the sample FRP material is not expected. The sleeping berth and double seat fires resulted in thermal exposure levels similar to that experienced in the cone calorimeter at a heat flux of  $50 \text{ kW/m}^2$ .

This approach can be helpful in establishing threat assessment and mitigation measures. For example, if luggage is expected in the vehicle, one can estimate how large a luggage fire might become. If one then can test a sample of the interior lining material in the vehicle, an assessment of the ability to ignite the interior lining can also be made. If so, further assessment of the likelihood for the interior material to spread the fire can be assessed, as discussed below. This data can lead to decisions regarding amounts of materials allowed in compartments, screening or detection technologies (fire detection, video monitoring, and so forth), or changes to the vehicle (e.g., changing or modifying materials or reducing the amount of combustible materials).

## SCREENING TOOL OF FLAME SPREAD

Armed with the knowledge of the range of potential initiation fires one might encounter and whether the initiation fires can ignite interior finishes and other contents, the next step is to assess whether interior materials might self-propagate the fire, leading to the potential for a fully involved fire encompassing the entire interior of the vehicle. A first-order screening approach has been established, making use of the  $b$ -parameter, to help evaluate if different vehicle lining materials are likely to support flame spread at various heat flux levels.

The  $b$ -parameter is a flame spread parameter that indicates a material's tendency to accelerate or decelerate flame spread when exposed to various heat fluxes. For any given heat flux, if the  $b$ -parameter is greater than zero, then flame is predicted to accelerate and spread; if the  $b$ -parameter is less than zero, the flame is predicted to decelerate and eventually extinguish itself. Assessing the  $b$ -parameter at various heat fluxes is important to properly characterize flame spread potential under different initiation fire and compartment fire conditions.

The  $b$ -parameter, as reflected in Equation 2, uses fire parameters measured or derived from cone calorimeter results with the variables explained below. The theoretical basis of this equation can be found in the work of Cleary and Quintiere (12).

$$b = 0.01\dot{E}'' - 1 - \frac{t_{\text{ig}}}{t_{\text{end}}} \quad (2)$$

where

$\dot{E}''$  = average HRRPUA ( $\text{kW/m}^2$ ),

$t_{\text{ig}}$  = time to ignition in the cone calorimeter, and

$t_{\text{end}}$  = end measured time in the analysis.

For the purposes of this study,  $t_{\text{end}}$  is measured as the time to flame out, that is, the time when HRRPUA is less than  $50 \text{ kW/m}^2$ .

HRRPUA is important as it is a measure of how large a fire might become given the fuels involved. HRRPUA values are derived from cone calorimeter results, and an average HRRPUA is then determined for use in the  $b$ -parameter analysis. Figure 3 shows derived HRRPUA

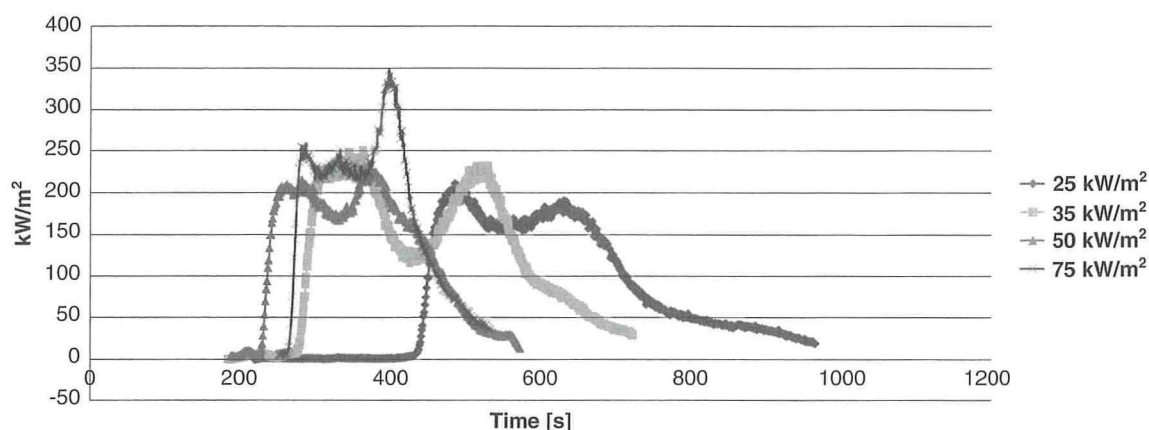


FIGURE 3 FRP gel coat cone calorimeter results for HRRPUA.

values based on test results for a common FRP material with a gel coat surface. FRP was selected for this analysis because FRPs are commonly used within many passenger rail vehicles as wall, ceiling, and in some cases seating material. HRRPUA data were collected over a range of incident heat flux levels to correspond to the range of initiation fires that might be expected within a vehicle.

An HRRPUA graph for each tested heat flux is then needed to develop HRRPUA averages for various time scales as well as to depict the appropriate time scale or length of time for each test, as geometric instabilities, cellular flames, and extended tail-like results can affect the analysis. Cellular flames, which are physically observed and recorded during each cone calorimeter test, often represent unstable flames that can flash during tests. Long tail-like results at the end of a cone test caused by nonuniform or edge burning yield extended values and times. This tail-like behavior can lead to lower HRRPUA averages if not addressed.

Figure 4 shows HRRPUA values throughout one test with various events noted. The time between ignition and the point when HRRPUA becomes less than 50 kW/m<sup>2</sup> results in averages that better represent the combustion behavior of the material.

For each of these ending times, an average HRRPUA is calculated by integrating the area underneath the curve. After the average HRRPUA,  $t_{ig}$ , and  $t_{end}$  are derived or observed from tests, the

$b$ -parameter value is calculated. Table 2 shows  $b$ -parameter values for FRP gel coat at heat fluxes of 25, 35, 50, and 75 kW/m<sup>2</sup>. The end of the experiment ( $t_{bo}$ , or time to burnout) was considered to be the time when the flame became cellular, actual flameout occurred, or the HRRPUA became less than 50 kW/m<sup>2</sup>. This is explained in more detail elsewhere (3, 4).

Whereas the  $b$ -parameter predicts the tendency of flame spread to accelerate or decelerate, results such as those presented in Table 2 can be used as a first-analysis screening tool for identifying a material's likelihood to spread flame in an acceleratory fashion. Figure 5 reflects a  $b$ -parameter data set developed for various materials (3).

Plotting materials on the same graph, as illustrated in Figure 5, gives a clear indication of which materials are predicted to accelerate or decelerate flame spread and at what radiant heat flux levels, which together provide critical insight into the potential for self-propagating flame spread (flame spread continues even after the initiation fire extinguishes) and resultant fire hazard (e.g., the fire stops or continues to spread and potentially involves the entire vehicle).

In Figure 5, the  $b$ -parameter is reflected on the y-axis, and the radiant heat flux level is shown along the x-axis. Data points reflect the calculated  $b$ -parameter at different heat flux levels. For example, the boxes reflect a glass-reinforced polymer (GRP) with a white gel coat. At heat flux levels of 10 and 20 kW/m<sup>2</sup>, the  $b$ -parameter is

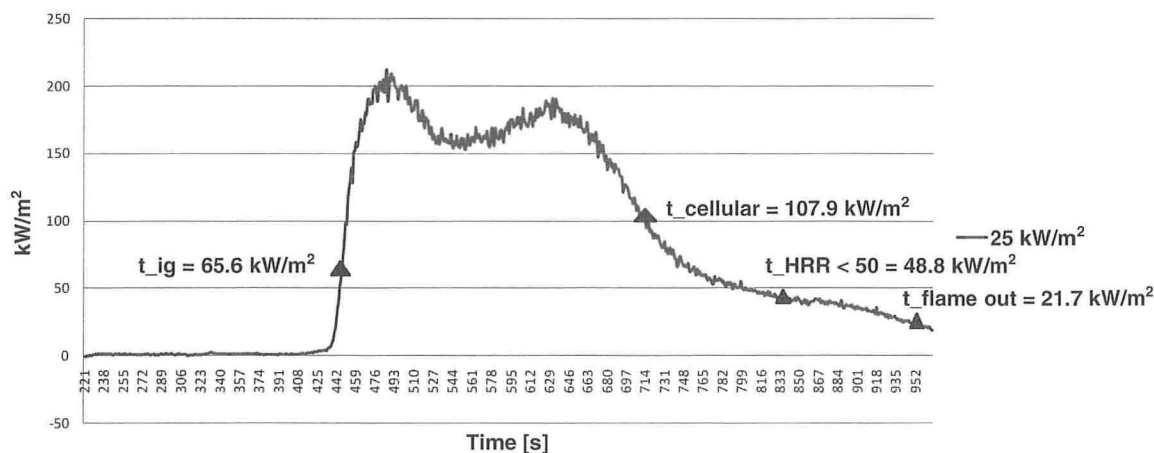


FIGURE 4 HRRPUA of FRP gel coat at cone calorimeter incident flux of 25 kW/m<sup>2</sup>.

TABLE 2 *b*-Parameter Values for FRP Gel Coat at Different Flux Levels

HRRPUA (kW/m <sup>2</sup> )	<i>b</i> -Parameter at End of Experiment		
	$t_{\text{cellular}} = t_{\text{bo}}$	$t_{\text{flame out}} = t_{\text{bo}}$	$t_{\text{HRRPUA} < 50} = t_{\text{bo}}$
25	0.041	-0.39	-0.15
35	0.35	0.034	0.14
50	0.30	-0.030	0.19
75	0.80	0.30	0.33

negative, indicating that the material is not expected to support self-propagating flame spread in the absence of an initiation fire or other heat source. However, at 50 and 70 kW/m<sup>2</sup>, the *b*-parameter is positive, indicating that once ignited, the material would continue to burn, as the resulting fire would provide sufficient energy to support accelerating flame spread. A brief discussion of this subject is presented below and is available elsewhere (3; Meacham et al., unpublished work). At a heat flux level of 25 kW/m<sup>2</sup> the *b*-parameter is near 0, indicating flame spread may or may not occur and that further analysis is needed (given the uncertainties in the material properties, testing, and prediction). Referring back to Table 1 and the in-text table above, a heat flux level of 25 kW/m<sup>2</sup> could be expected from a small trash fire. In these tests, the GRP with a cream-colored gel coat shows acceleratory tendencies under all heat flux levels, indicating ignition and self-propagation of the flame front from very small fires.

## COMBINING INITIATION FIRE, MATERIALS, AND INITIAL FLAME SPREAD

With the *b*-parameter providing an indication of the propensity for a material to propagate flame spread, a simplified upward flame spread model based on the work of Mowrer and Williamson (13) was developed to represent the initial flame spread that might occur in passenger rail vehicles (3; Meacham et al., unpublished work). Although the details of the flame spread model are not discussed here, it is helpful to understand how the initiation fire, its thermal threat, and the material properties fit together as part of the overall fire spread assessment process. Details regarding the flame spread model, how the outcomes of the model can be incorporated into a CFD model to assess how large a passenger rail vehicle fire might grow, and the potential impacts of that fire on life safety and critical infrastructure are presented elsewhere (3; Meacham et al., unpublished work).

In brief, upward flame spread is a complex phenomenon that involves thermal exposure from an initiation fire to a surface (e.g., a wall), sufficient heating of the surface material to release combustible gases (pyrolysis) and ignite them, and the ability of the material to continue to burn in the absence of the initiation fire by creating a sufficiently large and hot flame to continue the pyrolysis process.

Figure 6 illustrates the characteristics of upward flame spread as adapted from Mowrer and Williamson (13). The overall flame height ( $x_f$ ) consists of the flame created on the wall by the external source fire and the flame extension up the wall. Pyrolysis height is represented by  $x_p$ . When the fuel is considered spent or used up, it can no longer support a flame, and a burnout front develops, indicated by  $x_b$ . Physically, the initiation fire imposes an external heat flux on the vertical wall ( $q_e''$ ). After the first exposed material is ignited, the flame extends up the wall, emitting a flame heat flux ( $q_f''$ )

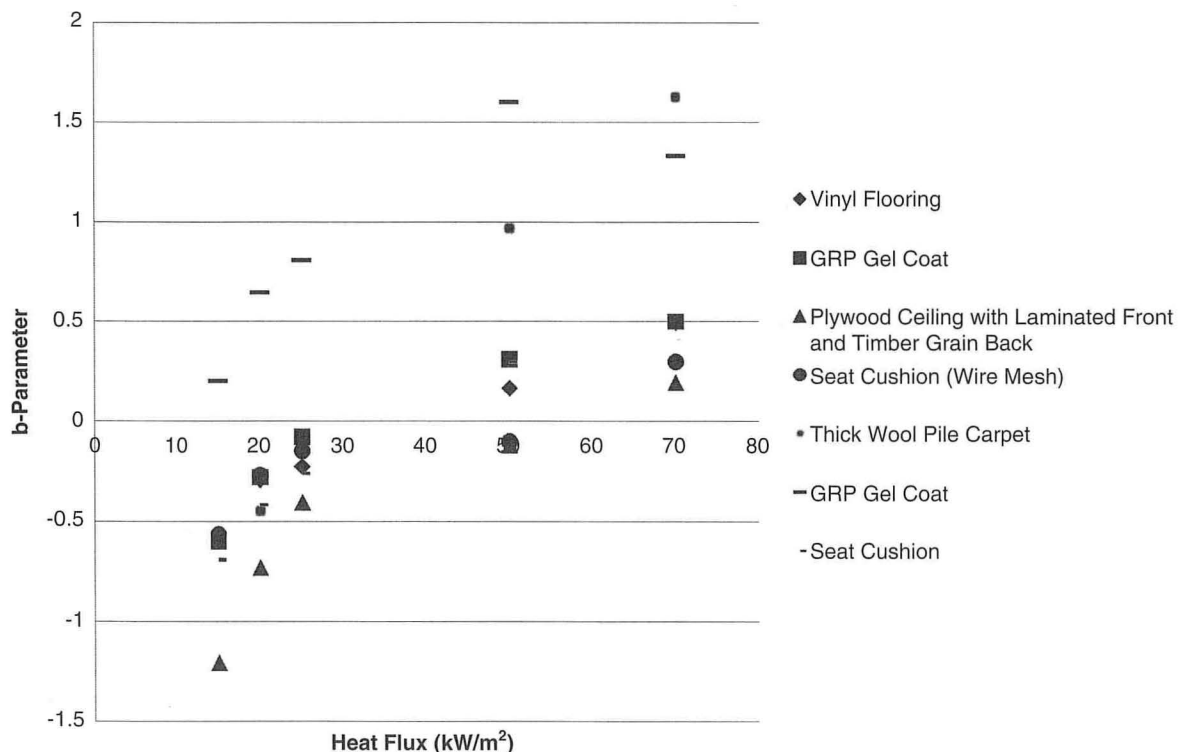


FIGURE 5 *b*-parameter data set (3).

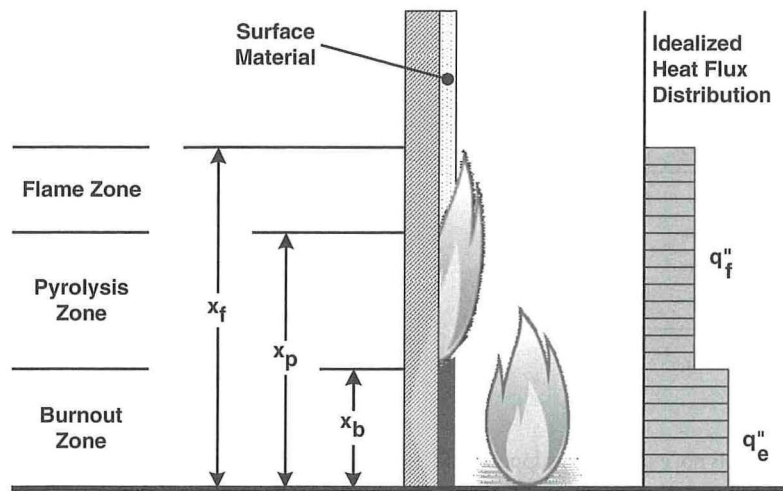


FIGURE 6 Model parameters of upward flame spread.

to the virgin fuel above. The idealized heat flux distribution reveals the potential heat fluxes from the initiation fire and the flame extension. In essence, the flame spread model uses the  $b$ -parameter and heat flux levels to assess the flame zone ( $x_f$ ) and pyrolysis zone ( $x_p$ ) to determine if sufficient burning will occur, in the absence of the initiation fire, to have the pyrolysis zone accelerate faster than the burnout zone ( $x_b$ ), meaning sufficient material will exist to self-propagate flame spread (3; Meacham et al., unpublished work).

The net result of this approach is a simplified means of assessing the potential for having a very large fire in a passenger rail vehicle by obtaining the fire properties of materials used in vehicles through small-scale cone calorimeter tests, applying the  $b$ -parameter screening tool, and assessing the potential for self-propagating flame spread, which could lead to full-vehicle burning. This process is accomplished by the following steps:

1. Review fire loss history and threat assessment data to understand types of fire scenarios of concern.
2. Characterize the initiation fires in terms of radiant energy and burning duration.
3. Obtain samples of interior materials and collect HRRPUA data by using the cone calorimeter over a range of heat flux levels.
4. Calculate the  $b$ -parameter for materials for different heat flux levels.
5. Develop a matrix of initiation fires that could lead to self-propagating fires, and thus to full-vehicle burning, for the materials in use.
6. Accept the risk associated with the expected initiation fires and existing materials, or take steps to prevent or mitigate the risk.
7. If self-propagating flame spread is indicated, employ the flame spread model and CFD analysis to assess the total fire size expected and potential impacts on life safety and infrastructure.
8. Take appropriate prevention, mitigation, or operational measures to control the fire risk.

## LIMITATIONS

This approach uses a combination of cone calorimeter test data and simplified analysis tools ( $b$ -parameter calculation and simple flame spread model). There is uncertainty associated with these components.

At the small-scale test level, differences in materials from one sample to another, variations in data collection from different test apparatus, and potential errors in reporting may exist. Development of common protocols for testing, apparatus calibration, and data reporting would help to reduce uncertainty in the data. Because the simplified analysis techniques are intended to provide a simple screening approach, some of the complexity associated with factors such as vehicle geometries was ignored (e.g., it was assumed that walls and ceilings were flat with no obstructions or curvature). Sensitivity analysis can be undertaken to better understand the importance of these issues to the level of accuracy needed in the screening approach.

## SUMMARY AND POTENTIAL APPLICATIONS

An approach has been developed that illustrates how small-scale test data from the cone calorimeter combined with a simplified screening approach and flame spread model can be used to assess interior finish materials in passenger rail vehicles for their propensity to spread fire given a range of initiation fires. The suggested approach offers a first-order understanding of fire risk (i.e., given the rail stock, what size initiation fires could result in self-propagating fires that could engulf the car) and requires only an understanding of the types of fires experienced within a particular rail system along with intentional fire scenarios of concern, characterization of those fires in terms of radiant energy and burning time, and estimation of  $b$ -parameters for interior finish materials, which can be developed from cone calorimeter tests. If more detail is needed (such as understanding the potential size of the resulting fire, how long it will take to develop, and what the potential impacts on life safety and critical infrastructure might be), the  $b$ -parameter and initiation fire data can be used along with a simplified flame spread model and CFD analysis to understand the fire in greater detail. The approach builds on recommended changes to current regulations (4) and has several potential applications:

- Fire hazard assessment of existing stock. Because it is costly to burn complete vehicles to understand fire development issues, this approach allows for small samples of materials to be tested (from an existing vehicle or from parts for a vehicle) and analysis to be conducted for an indication of what size initiation fire might lead to full-vehicle involvement and what the resultant fire size would be.

This assessment could be used as part of future TRVA of vehicles to understand impacts on passengers and other vehicles.

- Options analysis for new vehicle design. At the vehicle design stage, the methodology can be applied to help assess the performance of interior lining materials with respect to resistance to initiation fires of concern and contribution to overall fire size. Manufacturers can use this information to make informed judgments about material use in vehicles and cost-effective assessment of a wide range of material selection and configuration options.

- Fire hazard assessment of critical infrastructure. This methodology can increase understanding of the resultant fire size from a fully involved vehicle. This assessment could be used as part of a TRVA of stations, tunnels, and other rail infrastructure to assess impacts from the vehicle fire, from smoke control in a station to the thermal impact on tunnel linings or bridge decking. The methodology could be applied to existing infrastructure or new designs. With the current focus on high-speed rail, for example, such an approach may be beneficial for assessing a wide range of potential fire scenarios—accidental or deliberate—and to support assessment of mitigation options.

- Regulatory support. Current regulations generally require testing of material flame spread at a single heat flux level (1, 2, 4). However, as demonstrated by this research, a material may perform differently at different heat flux levels. At lower heat flux levels, a material might resist ignition and limit self-propagating flame spread, while at higher heat flux levels, the same material might readily ignite and support self-propagating flame spread. On the basis of level of risk or hazard deemed tolerable from a regulatory perspective, this methodology could be used to support modifying flame spread test requirements, such as perhaps requiring a range of incident heat fluxes and reporting outcomes rather than relying on pass–fail assessment at a single incident heat flux value (1). As new materials for rail vehicles are developed, application of these concepts could become part of regulatory benchmarking, helping to inform regulatory changes as deemed necessary and appropriate.

## ACKNOWLEDGMENT

This material is based on work supported by the Science and Technology Directorate of the U.S. Department of Homeland Security.

## REFERENCES

1. FRA. *Test Methods and Performance Criteria for the Flammability and Smoke Emission Characteristics of Materials Used in Passenger Cars and Locomotive Cabs*. 49 CFR 238, Appendix B. U.S. Department of Transportation, 2002.
2. FTA. *Hazard Analysis Guidelines for Transit Projects*. DOT-FTA-MA-26-5005-00-01. U.S. Department of Transportation, 2000.
3. Meacham, B. J., N. A. Dembsey, K. Schebel, M. Johann, and J. Tubbs. *Rail Vehicle Fire Hazard Guidance*. Final report. International Cooperative Programs Office, Science and Technology Directorate, Department of Homeland Security, Washington, D.C., 2010.
4. National Association of State Fire Marshals. *Recommended Fire Safety Practices for Rail Transit Materials Selection*. Project No. DC-26-5243-00. Prepared for FTA, U.S. Department of Transportation, 2008.
5. FTA, U.S. Department of Transportation. *Rail Safety Statistics Report*. [http://transit-safety.fta.dot.gov/publications/RailSafety/Rail\\_Safety\\_Statistics\\_Report\\_2009-FINAL.pdf](http://transit-safety.fta.dot.gov/publications/RailSafety/Rail_Safety_Statistics_Report_2009-FINAL.pdf).
6. Babrauskas, V. Heat Release Rates. In *SFPE Handbook of Fire Protection Engineering* (P. DiNenno, ed.), 4th ed. NFPA, Quincy, Mass., 2008, pp. 3-1–3-59.
7. Babrauskas, V. The Cone Calorimeter. In *SFPE Handbook of Fire Protection Engineering* (P. DiNenno, ed.), 4th ed. NFPA, Quincy, Mass., 2008, pp. 3-90–3-108.
8. *Standard Test Method for Heat and Visible Smoke Release Rates for Materials and Products Using an Oxygen Consumption Calorimeter*. ASTM E-1354. *Annual Book of ASTM Standards*, ASTM International, West Conshohocken, Pa., 1997.
9. *Reaction to Fire Tests: Heat Release, Smoke Production, and Mass Loss Rate Standard*. ISO 5660-2. International Organization for Standardization, Geneva, 2002.
10. Huggett, C. Estimation of the Rate of Heat Release by Means of Oxygen Consumption. *Fire and Materials*, Vol. 12, 1980, pp. 61–65.
11. Janssens, M. Calorimetry. In *SFPE Handbook of Fire Protection Engineering* (P. DiNenno, ed.), 4th ed. NFPA, Quincy, Mass., 2008, pp. 3-60–3-89.
12. Cleary, T., and J. Quintiere. A Framework for Utilizing Fire Property Tests. *Proc., 3rd International Symposium on Fire Safety Science*, Elsevier Science Publishers Ltd., London, 1991, pp. 647–656.
13. Mowrer, F. W., and R. B. Williamson. Flame Spread Evaluation for Thin Interior Finish Materials. *Proc., 3rd International Symposium on Fire Safety Science*, Elsevier Science Publishers Ltd., London, 1991, pp. 689–698.

---

*The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.*

*The Passenger Rail Equipment and Systems Integration Committee peer-reviewed this paper.*

# Geometry of High-Speed Turnouts

Miguel R. Bugarín, Alfonso Orro, and Margarita Novales

Turnouts are singular points of the railway track. In the past 20 years, a series of advances has added to their design as well as to the design of other elements of the track structure. These developments have allowed vehicles both to increase their running speed over the turnouts and to improve their reliability and security, and thus reduce maintenance costs. This paper focuses on geometric improvements in turnouts that permit high speeds over direct (350 km/h) and diverging (160 to 220 km/h) tracks. These improvements are related to diverging track alignments. The improvements include transition curves, switch rail design, and mechanization that have been adopted to avoid the straight switch rail strike phenomenon, and they include crossing modifications that avoid the existence of the gap.

A railroad turnout is a mechanical installation that enables trains to be guided from one track to another at a railway junction. These elements of the railroad track have always been singular points of great complexity as a result of one of the main characteristics of railways: the automatic guidance of vehicles by means of conicity and the use of interior flanges in the wheels, which link the vehicle path to the track layout. The flange requires appropriate space for the flange (flange path) to move forward without finding obstacles; this requirement leads to the particular components of the turnout that are discussed in the next section (1).

Since the first mine turnouts were designed by John Curr in 1976, these elements have constituted singular points in the track, with expensive inspection and maintenance tasks. Indeed, Swiss Railways estimates that between 20% and 40% of maintenance expenses are related to the inspection, maintenance, and renewal operations of turnouts and crossings (2). The maintenance cost of a turnout is equivalent to that of 300 to 500 m of conventional track.

So it is not surprising that the design evolution of turnouts has been oriented to improve the safety and speed of vehicles when running over them, reducing their life-cycle costs and extending their lifespan. Because these objectives are especially critical in turnouts for high-speed lines, this type of turnout uses the most modern technology and the most advanced solutions, which make getting running speeds of up to 220 km/h over the diverging track possible while guaranteeing the safety and reliability of the system.

## CRITICAL POINTS OF CONVENTIONAL TURNOUTS IN RELATION TO SPEED

Figure 1 shows the geometry of a simple right-hand turnout for a conventional (not high-speed) track with its main elements.

ETS Ingenieros de Caminos, Canales y Puertos, La Coruña University, Campus de Elviña, s/n, E-15071, Corunna, Spain. Corresponding author: M. Novales, mnovales@udc.es.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 64–72.  
DOI: 10.3141/2261-08

Switch rails guide the vehicle over the direct or the diverging track, according to the position of each switch rail in relation to its corresponding stock rail. In Figure 1, the positions of the curved and straight switch rails cause the turnout to guide the vehicle over the diverging track. If these elements are in the opposite position, the turnout guides the vehicle toward the direct track.

The crossing is the turnout part that allows the two routes to cross each other while avoiding any obstacle that the vehicles' flanges may encounter in their movement. For this to happen, an empty space known as the gap must be left between the two rails that cross each other in the crossing.

To guarantee the support of the wheel that is running over the crossing despite the existence of the gap, it is necessary to enlarge the connection rails to both sides of the frog. These parts of the crossing are known as wing rails. However, the gap causes the wheel running over the frog to lose its guidance. For example, if the vehicle is running over the turnout toward the diverging track (as shown in Figure 1), it will be guided by the contact of the left wheel flange with the curved switch rail, but when it reaches the gap, this wheel would lose its guidance because of the empty space; as a result, it could strike the frog nose or even go toward the direction for which the turnout is not provided (in this case, toward the direct track). To prevent this from happening, check rails are disposed in the crossing in such a way that when the left wheel is located over the gap, the right wheel can guide the vehicle by means of the contact between its interior face and the check rail. In this way, guidance over the crossing is also guaranteed.

In a conventional turnout with the characteristics described above, the critical points in relation to running speed are the following:

- A discontinuity in the course when vehicles run toward a diverging track, usually without the existence of a transition curve. This phenomenon leads to a discontinuity in the transversal acceleration or overacceleration caused by the impossibility of establishing the appropriate superelevation for the curve.
- A discontinuity in the track stiffness, notably at the level of the crossing, both vertical and horizontal.
- Finally, and in particular, a discontinuity in the support of the wheels when they pass through the crossing, when the wheel moves from the wing rail to the tip of the frog, or vice versa, crossing the gap (3).

In the following sections specific characteristics of high-speed turnouts are discussed that counteract the effects of these critical points, using as examples the Spanish high-speed turnouts of the Madrid–Seville and Madrid–Barcelona lines.

## TURNOUT GEOMETRIC ALIGNMENT

As railroad track elements, turnouts must comply with specific operational requirements. Railway administrations distinguish among turnouts by radius size. Turnouts suitable for use on a main track

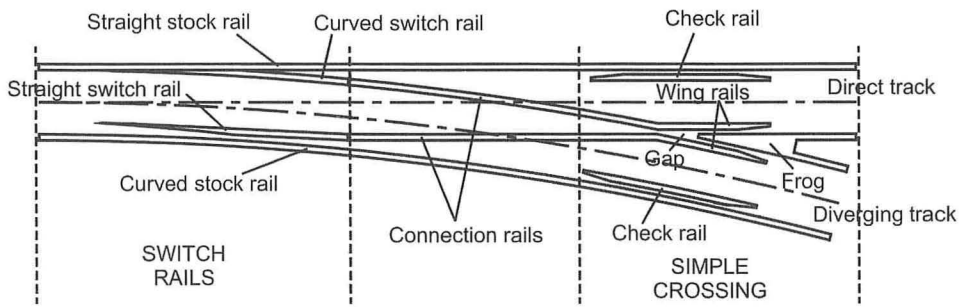


FIGURE 1 Geometry of simple right-hand turnout for conventional tracks.

have a large diverging track radius and low values of the crossing angle tangent. Turnouts with a smaller diverging track radius and a higher angle are used in railroad yards or stations, leading to a better use of the land surface available. In other words, the use of a turnout is determined by its main parameters: diverging track radius ( $R$ ) and crossing angle tangent ( $1:n$ ).

The element that determines the diverging track radius for a high-speed turnout is the maximum running speed over that track. The relation between radius and speed is given by three fundamental parameters that are enforced by each railroad administration according to its experiences and criteria. These parameters are

- Unbalanced centrifugal acceleration over diverging track,
- Overacceleration, and
- Jerk.

### Centrifugal Acceleration

The centrifugal acceleration experienced by a body that travels over a circular curve of radius  $R$  (in meters) at a speed of  $v$  (in meters per second) is

$$a_c = \frac{v^2}{R} \quad (1)$$

where  $a_c$  is centrifugal acceleration. In a circular curve without superelevation (the case of the diverging track in a turnout provided in a straight direct track), the value of the unbalanced centrifugal acceleration ( $a_{cub}$ ) is the same as the centrifugal acceleration ( $a_c = a_{cub}$ ). So, Equation 1 can be used to determine  $a_{cub}$ .

It is common to transform Equation 1 to introduce speed in kilometers per hour and obtain

$$\left. \begin{aligned} a_{cub} &= \frac{v^2}{R} \\ v_{(m/s)} &= \frac{V_{km/h}}{3.6} \end{aligned} \right\} \rightarrow a_{cub} = \frac{V^2}{12.96 \cdot R} \quad (2)$$

With Equation 2 the maximum speed allowed over the diverging track ( $V$ , in kilometers per hour) can be determined, imposing the condition that the vehicle does not surpass the maximum value of unbalanced centrifugal acceleration. In this way, if the limit value of  $a_{cub}$  is  $0.65 \text{ m/s}^2$  (a common limit), the relation between maximum speed  $V$  ( $V_{max}$ ) and circular radius  $R$  of the diverging track will be

$$V_{max} = 2.9 \cdot \sqrt{R} \quad (3)$$

In turnouts with a curved direct track, the value of superelevation should be considered. In any case, this value will be the same for the diverging and direct tracks because of construction requirements, and normally it will be adapted to the needs of vehicles running over the direct track.

### Overacceleration

Because the geometry of railroad tracks must allow a progressive transition between different alignments with different transversal accelerations in order to avoid sudden changes of values, it is necessary to provide a transition alignment between straight track and circular curves.

The variation of transversal acceleration with time  $t$  (its derivative) is known as overacceleration ( $\psi$ ) or instantaneous quality. This parameter is one of the criteria used to guarantee passenger comfort. Its mathematical expression, when there is no superelevation, is as follows:

$$\psi = \frac{da_{cub}}{dt} = \frac{d\left(\frac{V^2}{12.96 \cdot \rho}\right)}{dt} \quad (4)$$

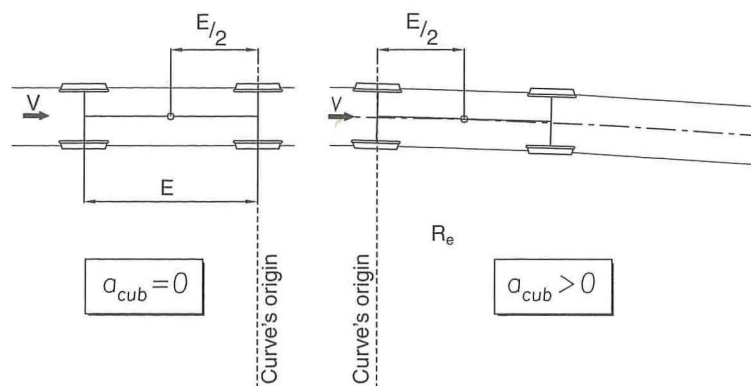
In a circular curve, the value of  $a_{cub}$  is constant, as is  $\rho$  (with value  $R$ ). Therefore, overacceleration will be nil. However, in a transition curve the radius ( $\rho$ ) changes. If the transition is, for example, a clothoid (Euler spiral), the overacceleration value will be constant and not nil through the transition curve.

Overacceleration is limited, mainly, because of its influence on comfort. Nonetheless, it is related to the appearance of supplementary transversal forces between the vehicle and the track, which was shown by studies performed around 1970 by the Track Equipment and Maintenance Institute of the former USSR (4).

### Jerk

When a vehicle, considered as a mass concentrated on one point, passes directly from a straight alignment to a circular curve, the unbalanced transversal acceleration value changes suddenly from zero to  $a_{cub}$ . This fact would imply an infinite theoretical overacceleration.

In real life the vehicle has a certain length, so it is not a mass concentrated on one point. Unbalanced transversal acceleration will not act with its whole value in the vehicle's center of mass until the whole length of the vehicle is in the circular curve, that is, until its last truck or wheel set is located in the curve's origin (Figure 2). At that moment, the vehicle will have moved the distance between its wheel sets or trucks (wheel base or bogie base of the vehicle) forward.

FIGURE 2 Interval between 0 and maximum values of  $a_{cub}$ .

In this case, the concept of overacceleration as derivative of unbalanced transversal acceleration with time loses its sense. The parameter of interest is now the average value of unbalanced transversal acceleration in the time used to get from the first to the second position in Figure 2 (5).

This value ( $\psi^*$ ) is known as “jerk” in English, *ruck* in German, *choc* in French, and *empellón* in Spanish [the latter proposed by García Díaz-de-Villegas (6)]. It is formulated in the following equation, where  $E$  is the distance shown in Figure 2, and  $v$  is the vehicle’s velocity:

$$\psi^* = \frac{\Delta a_{cub}}{\Delta t} = \frac{\Delta a_{cub}}{\frac{E}{v}} \quad (5)$$

The influence of jerk on comfort, track geometry degradation, and other subjects is not yet well known. In fact, this is an open question for research. Some interesting contributions to date include the following:

- Theoretical works by Megiery, who makes a thorough vectorial treatment of railroad alignment, suggesting layouts with continuous function and derivative (7), and
- Contributions by Nasarre et al. with a new interpretation of jerk (8).

### Quantification of Parameters

Railroad administrations quantify the dynamic parameters that determine track geometry. Some administrations distinguish between recommended values for broad tracks and for turnouts, although most bear in mind the total continuity criterion and do not establish less strict recommendations for turnouts as they are a part of the whole track.

For Spanish high-speed turnouts, the values shown in Table 1 have been set.

### Evolution of Turnout Geometry to Comply with High-Speed Requirements

From the information above it is easily deduced that simple circular curve layouts must be abandoned as the running speed over the diverging track of the turnout increases, and other kinds of alignments must be used that reduce overacceleration and jerk values. This is why layouts based on two circular curves, which were first used in German turnouts in the 1980s, have been rejected since the 1990s, giving way to layouts with clothoids or cubic parabolas.

In the Paris Sud-Est high-speed line, the French national railroad (SNCF) introduced turnouts with a diverging track consisting of a transition curve (cubic parabola) tangent to the straight alignment with an initial radius of 3,000 or 6,720 m (because it is physically impossible to provide a parabolic transition starting from an infinite radius on a switch rail). This parabola ends with an infinite radius at the frog nose (9).

Swiss railroads (SBB) were the first to introduce turnouts with a diverging track consisting of two clothoid arcs. This solution, known as a tip clothoid, allows a reduction in the switch’s deviation angle with respect to the French geometry. This is the geometry currently proposed by most German manufacturers, with a configuration known as asymmetrical clothoids that consists of two clothoids with different parameters linked with each other.

The tip clothoid solution, however, has the inconvenience of generating high transversal acceleration values when trains run over the turnout at high speeds. For this reason, the working group constituted to develop the design for high-speed turnouts for the Madrid–Seville line (with experts from Renfe, TIFSA, and Cantabria University) set out a new kind of layout known as the plateau clothoid. This solution uses two clothoid arcs connected by a circular curve (6). The greater turnout performance this geometry provides for Spanish high-speed lines from Madrid to Seville and to Barcelona is presented schematically in Figure 3.

TABLE 1 Limit Values of  $a_{cub}$ , Overacceleration, and Jerk in Spanish High-Speed Turnouts

Turnout	$a_{cub}$ (m/s <sup>2</sup> )	$\psi$ (m/s <sup>3</sup> )	$\psi^*$ (m/s <sup>3</sup> )
DSIH-60-10000/4000-0.026-CM-D (Madrid–Seville high-speed line)	Normal: 0.51 Maximum: 0.65	Maximum: 0.40	Normal: 0.40 Maximum: 0.85 Exceptional: 1.20
DSIH-60-17000/7300-0.02-CM-D (Madrid–Barcelona high-speed line)	0.50	0.60	1.10

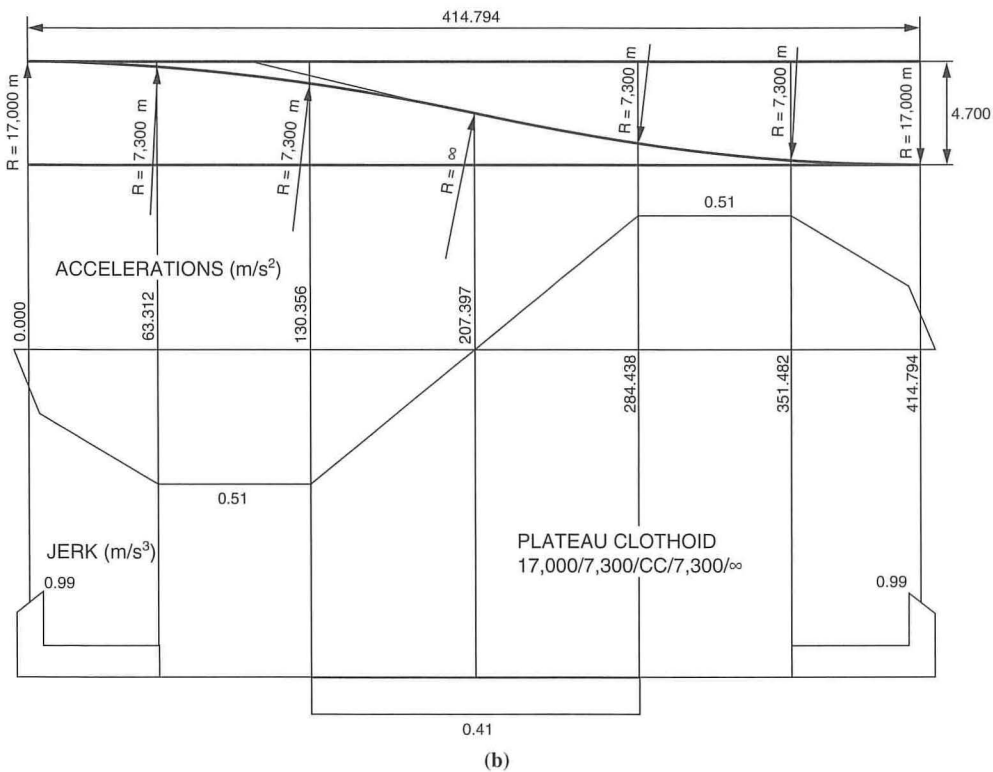
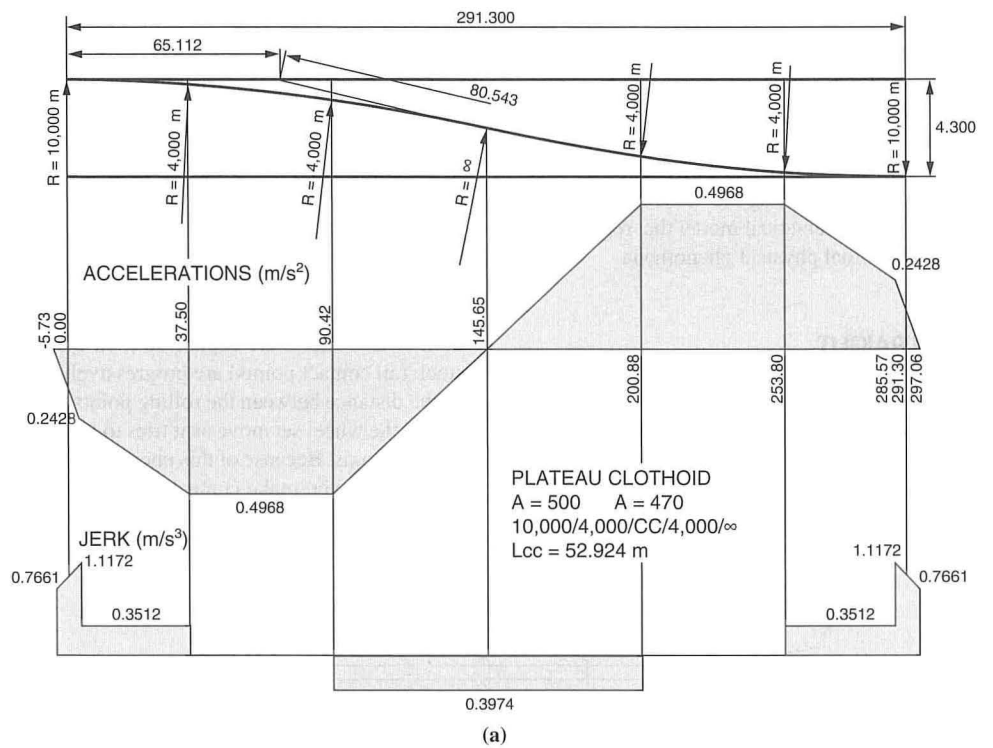


FIGURE 3 Turnouts: (a) DSIH-60-10000/4000-0.026-CM-D ( $V_{\max}$  over diverging track = 160 km/h) and (b) DSIH-60-17000/7300-0.02-CM-D ( $V_{\max}$  over diverging track = 220 km/h) ( $L_{cc}$  = length of circular curve).

## IMPROVEMENT IN SWITCH GEOMETRY

To obtain design improvements that reduce the interaction forces between the vehicle and the turnout, it is necessary to analyze the movement of railroad vehicles through switch and crossing, considering every situation that can affect vehicle running (e.g., mechanical characteristics of the vehicle, weight, speed, initial conditions when entering the switch) as well as turnout characteristics (e.g., geometry, kind of rails, worn profile, switch position). The first stages of such a design begin by simulating the vehicle's behavior over the turnout by means of a mathematical model that represents, with acceptable accuracy, the actual physical phenomena.

## PHENOMENON OF STRAIGHT SWITCH RAIL STRIKE

The conical shape of railroad wheels and the inherent wheel set property of making the wheels rotate at the same speed without differential turns between them have been a traditional way of guiding in railroads. This design, which has been used since the beginning

of the 19th century and is still used today even in high-speed trains, is the origin of a disturbing movement known as hunting. By means of this movement, the wheel set reacts to a specific disturbance that moves it away from its equilibrium position (centered in the track and with its rotational axis perpendicular to it) by trying to relocate its center of mass over the track center by means of a series of oscillations around it.

A change of track gage is one of the specific situations that results in a hunting movement. This phenomenon appears when the vehicle wheel sets run over a turnout switch advancing through the direct track. When the straight switch is not wide enough to maintain the supporting contact with the corresponding wheel, this wheel will have to roll over the curved stock rail. Therefore, while the wheel set is penetrating into the switch, the points of the rails over which both of the wheels are supported (i.e., the curved and the straight stock rail contact points) are progressively farther from each other, so the distance between the rolling points increases (Figure 4). This makes the wheel set move as it tries to locate its axis over the effective track axis. Because of this phenomenon, the flange of one of the wheels can even make contact with the straight switch rail, which will lead the vehicle to move along the direct track (10). This strike

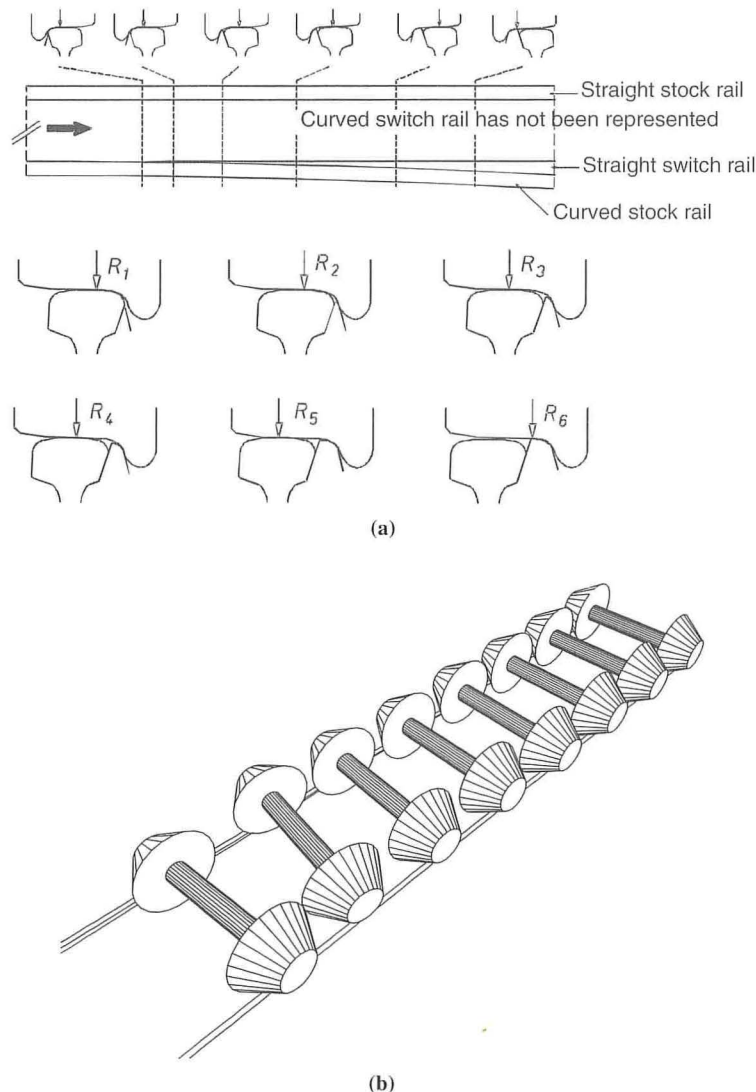


FIGURE 4 (a) Gage increase when wheel set is running through direct track and (b) induced straight switch rail strike phenomenon.

movement lasts until the wheel is supported by the straight switch. At that time, the track gage returns to its normal value, and the wheel set will try to center its axis in relation to the new track axis. This entire movement is diagrammed in Figure 4.

The strike movement of the wheel sets against the straight switch rail leads to

- Impacts and consequent wear in turnout elements, specifically in the straight switch rail, as well as in the vehicles' wheel sets and trucks;
- Higher maintenance costs, mainly in switch elements;
- Noise; and
- Discomfort.

Today there are two systems, the Fakop system and the CATFERSAN system, which avoid impacts against the straight switch rail; these are discussed below.

### Fakop System

Spanish high-speed lines are the first in the world to use the Fakop system in their turnouts. The Fakop system (an acronym for Fahrkanteoptimierung, or rail edge optimization) was developed by the German company Butzbacher Weichenbau GmbH.

The Fakop system induces an opposite movement to the one that makes the wheel set strike the straight stock rail. This opposite movement is achieved by enlarging the distance between both stock rails (straight and curved) in an adequate way and in a localized area. The enlargement is achieved by locally curving the straight stock rail, adopting the shape shown in Figure 5 (the scales have been altered so that the stock rail modification can be observed). With this straight stock rail geometry, at the same time that the wheel over the curved stock rail is experiencing a movement of its contact point toward the exterior, the wheel over the straight stock rail will be supported by a contact point that is separated more and more from the curved stock rail. In this way, the track axis of the direct track remains straight, and so the movement against the straight switch rail to center the wheel set axis over the track axis disappears, while the wheel set and the track already have the same axis (11).

Both the length and the geometry of the straight stock rail to be curved are obtained by trial and error and analysis of the results of a vehicle's simulated behavior when running over the modified switch of the turnout. Thus, a near-optimum solution rather than an optimum solution is achieved.

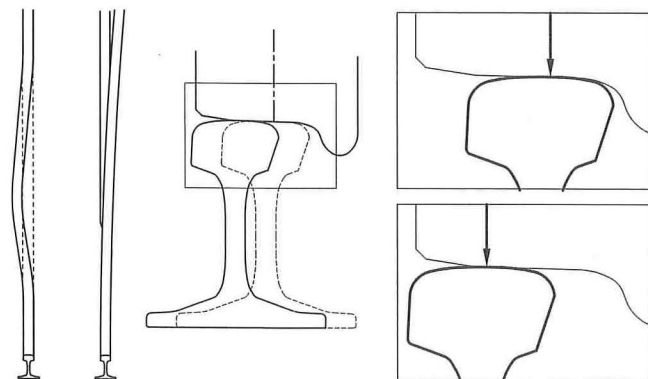


FIGURE 5 Fakop system.

The main drawbacks of this solution are related to the manufacturing process. Modifying the geometry of the straight stock rail leads to a more complex mechanization of the curved switch rail that has to fit it. In addition, it is necessary to calculate the position variation of the drill holes for fastenings in the turnout sleepers, bearing in mind concrete retraction. These considerations all lead to more difficult production and noticeable increases in design and construction costs. This solution also requires the adoption of specific techniques for tamping, aligning, and leveling operations in the switch zone in order to accurately maintain the desired modified geometry in the straight stock rail.

The main advantage of this solution consists of the improvement of the vehicle's behavior when running over both the direct and the diverging track. Kinematics gage optimization is a system based on the Fakop solution (12).

### CATFERSAN System

The CATFERSAN solution (8, 13) reprofiles the straight stock rail head in a specific zone with a precise geometry in such a way that contact between the wheel and stock rail is avoided in this area, but the flange guidance is retained (Figure 6).

Like the Fakop solution, in the CATFERSAN solution the length and geometry of the mechanized area are determined by a trial-and-error strategy, simulating the behavior of the most aggressive vehicle running over the switch zone by means of a railroad's dynamics software.

The mechanization of the stock rail head is effective for solving the strike phenomenon against the straight switch rail and obtaining, with reasonable costs, the following benefits:

- Decreased lateral forces over the straight switch, which lead to decreased lateral wear in the switch;
- Softer lateral movements of the railway vehicles, which lead to decreased actions over the rest of the elements of the turnout, thus lowering maintenance costs; and
- Improved passenger comfort.

The CATFERSAN solution does not require the design modification of turnout sleepers, nor are specific maintenance procedures needed for the turnout switch zone.

Figure 7 shows the wheel set lateral displacement obtained by simulation of the vehicle's behavior over a high-speed turnout prototype (developed by the Spanish firm Felguera Melt S.A.) that can

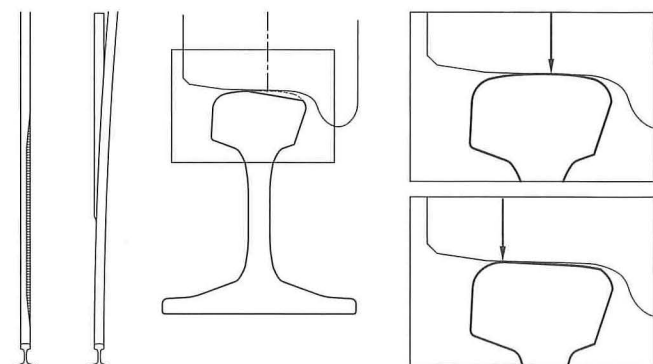


FIGURE 6 CATFERSAN system.

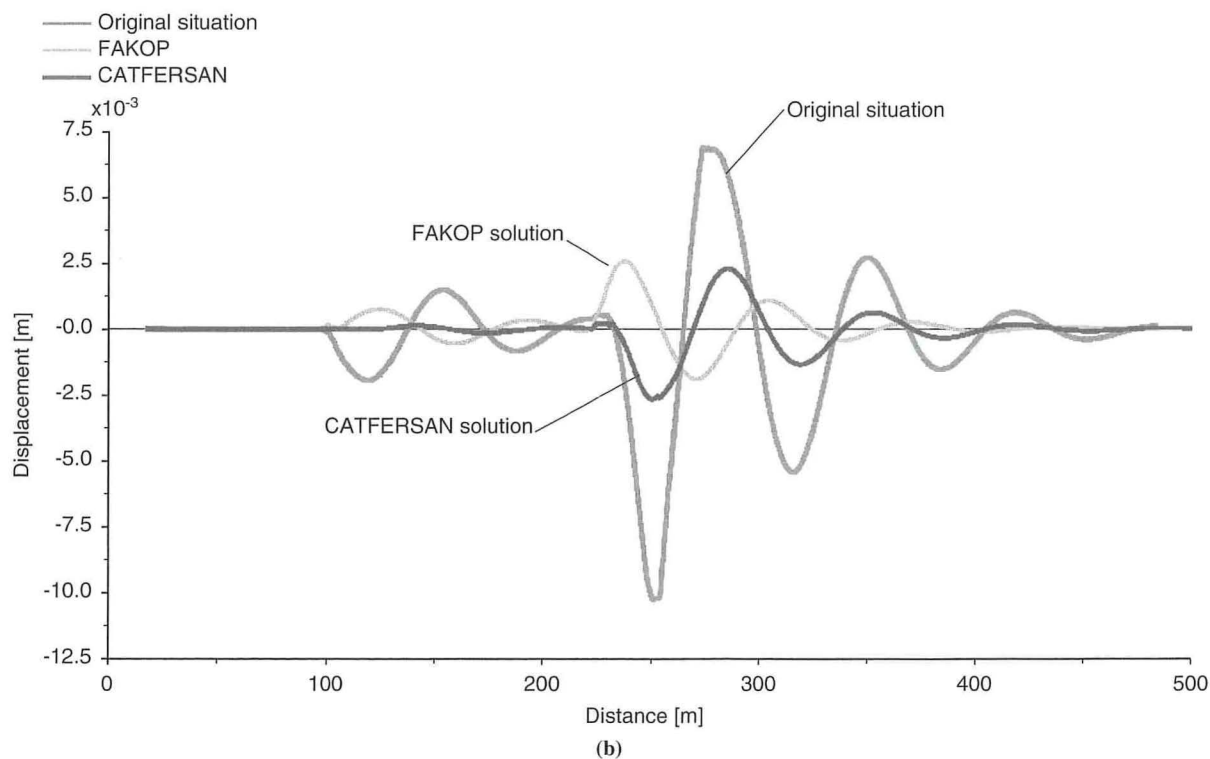
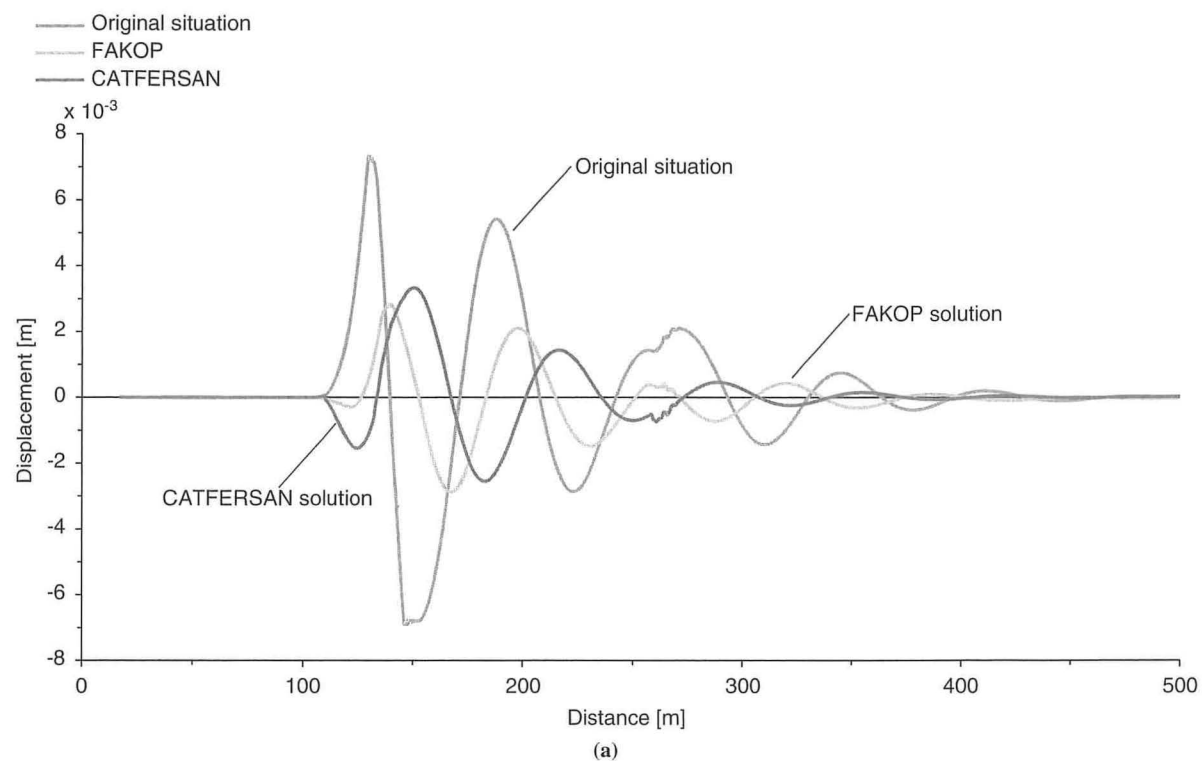


FIGURE 7 Lateral displacement of first wheel set when vehicle is running (a) from switch to frog and (b) from frog to switch (speed = 300 km/h; initial lateral displacement of wheel sets, trucks, and car body = 0; initial wheel set attack angle = 0).

be used for speeds up to 220 km/h over a diverging track with the Fakop and the CATFERSAN solutions. Figure 7a shows the lateral displacement of the first wheel set of the Spanish high-speed train AVE when running over the turnout from the switch to the frog, and Figure 7b shows the same displacement when running in the opposite direction.

## CROSSING IMPROVEMENT

To obtain a higher-speed circulation over the diverging track, it is necessary to increase the diverging track radius in such a way that limit values for centrifugal acceleration, overacceleration, and jerk are observed (see above). This increase of the radius value leads to a reduction of the crossing angle and, as a consequence, to a longer gap in the crossing zone. The increased gap allows an appropriate flange path for the flange to pass toward both the diverging and direct tracks. If this gap enlargement is excessive, however, it could prevent the safe running of vehicles over the crossing. In this case, the gap must be removed.

The gap generates a vertical and lateral discontinuity that causes high impact forces, jerks, discomfort, and excessive wear. Because these problems increase at higher speeds, it is recommended to remove the gap for speeds over 160 km/h.

In order to remove the gap, movable elements are provided to keep an adequate distance between the frog and the wing rails (flange path) in the appropriate position relative to the switch rails (i.e., appropriate for the turnout given the circulation over the direct or the diverging track). In this way, check rails become unnecessary, although some railroad administrations keep them as an additional safety measure.

Although until 10 years ago movable wing rails were used to remove the gap, today all railroad administrations with high-speed lines use a movable frog nose that can be put together with the corresponding wing rail. In this way, the gap and one of the flange paths are removed, keeping the other path empty to allow the flange to pass through it.

The two ways of achieving a moveable frog nose, the German and the French designs, are discussed below.

### Movable Frog Nose of German Design

The movable frog nose provided in the Madrid–Seville high-speed line is the heavy standard type from the German national railroad (DB) (Figure 8). The central part of the elastic frog nose consists of a single piece forged homogeneously and manufactured from a standard steel section.

To allow the movement of the frog nose to fit the corresponding wing rail, a device is provided in the rail that avoids the constitution of an undeformable triangle. The movable frog nose is protected by being rounded and sheltered 3 mm under the wing rail (similar to the switch rail protection under the stock rail) in such a way that the wheel can only touch it tangentially. Connection [International Union of Railways (UIC) 60] rails are welded to the mechanized nose by means of flash welding.

The longitudinal forces of the continuously welded rails are transmitted from wing rails to movable nose backs by means of two or three pads screwed down by high-resistance screws and auto-blocking nuts. Other pads are provided between the two rails on the back of the movable nose, screwed down in the same way, to equalize the forces of the two rails (Figure 8).

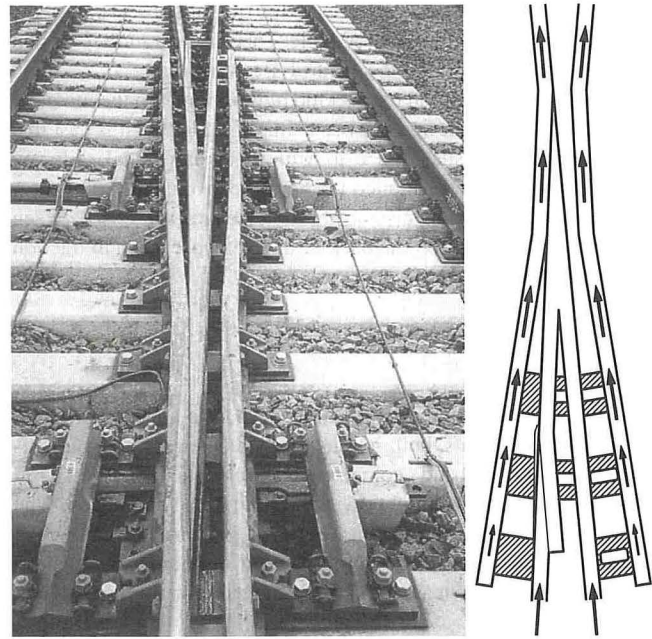


FIGURE 8 Movable frog nose of German design.

### Movable Frog Nose of French Design

In the movable frog nose turnouts provided in French high-speed lines (Figure 9), the two main elements of the frog are

- The berth, which plays, in relation to the movable frog nose, the same part as the stock rail in relation to the switch rail, and
- The elastic movable point.

The movable frog nose is obtained by mechanized rails with a UIC 61 A profile (Figure 9). It is composed of the point itself and the checkpoint, which are linked by studs. This point–checkpoint

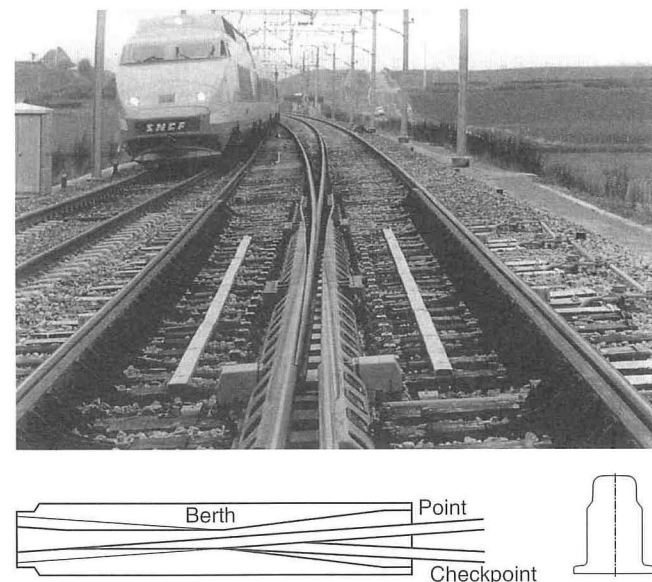


FIGURE 9 Movable frog nose of French design.

configuration allows the movement of the frog nose by avoiding the formation of an undeformable triangle. This configuration is used in the high-speed turnouts of the Madrid–Barcelona line.

In the latest conception of this kind of crossing with the movable frog nose, the berth design has been simplified so that it is cast in a single manganese–steel piece to which standard steel rails are welded by means of a patented method. This welded joint guarantees the continuous transmission of the welded rail's forces between the berth and the rails on the back of the frog nose, both in traction and in compression.

## CONCLUSION

This paper presents the modifications needed to allow high-speed running over turnouts, both in relation to geometric improvements in the turnout layout and geometric improvements in the switch rail design to avoid the straight switch rail strike phenomenon. The crossing design modifications necessary to prevent the existence of the gap are also described. Turnouts used in Spanish high-speed lines have been shown as examples of the modifications explained.

## REFERENCES

1. Bonaventura, C. S., A. M. Zarembski, J. W. Palese, and D. R. Holfeld. Increasing Speeds Through the Diverging Route of a Turnout Without Increasing Lead Length. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1836, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 61–67.
2. Zwanenburg, W.-J. Modelling Degradation Processes of Switches and Crossings for Maintenance and Renewal Planning on the Swiss Railway Network. Presented at 6th Swiss Transport Research Conference, Ascona, Switzerland, 2006.
3. Alfi, S., and S. Bruni. Mathematical Modelling of Train–Turnout Interaction. *Vehicle System Dynamics*, Vol. 47, No. 5, 2009, pp. 551–574.
4. Yershkov, O. P. Valeurs admissibles des accélérations non compensées et leur variation en fonction du temps sur les lignes ferroviaires à grand vitesse. *Rail International*, No. 5, 1979.
5. Oswald, J. R. Turnout Geometry Optimization with Dynamic Simulation of Track and Vehicle. VAE Aktiengesellschaft, Zeltweg, Austria, 2000.
6. García Díaz-de-Villegas, J.-M. Consideraciones sobre la geometría de los desvíos para Alta Velocidad. Presented at Jornadas Europeas de Alta Velocidad Ferroviaria, Madrid, Spain, 1988.
7. Megiery, J. *Eisenbahn-Bewegungsgeometrie*. Akadémiai Kiadó, Budapest, Hungary, 1993.
8. Nasarre, J., M. Cuadrado, P. González, and E. Romo. Analysis Method of the Ride Quality in Railway Alignment with Geometrical Singularities. World Congress on Railway Research, Cologne, Germany, 2001.
9. Oeconomos, J. Les appareils de voie de très petit angle de la Ligne à Grande Vitesse de Paris à Lyon. *Rail International*, No. 1, 1983.
10. Kassa, E., J. Perez, S. Iwnicki, and P. Allen. Optimization of Track Gauge and Track Stiffness Along a Turnout Using a Multibody Simulation Tool. *Proc., 21st International Symposium: Dynamics of Vehicles on Roads and Tracks IAVSD*, Stockholm, Sweden, 2009.
11. Bugarín, M. R., and J.-M. García Díaz-de-Villegas. Improvements in Railway Switches. *Proc., Institution of Mechanical Engineers, Journal of Rail and Rapid Transit*, Vol. 216, No. 4, 2002.
12. Budisa, M. Selecting a Turnout Design. *Railway Track and Structures*, Vol. 104, No. 7, 2008.
13. García Díaz-de-Villegas, J.-M., and M. Rodríguez Bugarín. *Desvíos Ferroviarios*. Editorial Ingenieria Cantabria, Santander, Spain, 1995.

*The Railroad Track Structure System Design Committee peer-reviewed this paper.*

# International Variation in Cost–Benefit Analysis of Urban Rail Projects

## Impact on Outcomes

Evan Gwee, Graham Currie, and John Stanley

This paper compares urban rail cost–benefit analysis (CBA) in 11 countries and the region of Hong Kong, China, to understand differences and implications for evaluation outcomes and to highlight possible CBA improvements. All countries studied used multicriteria analysis and examined capital, operating, and maintenance costs, and some included asset residual value. Monetized benefits varied, although travel time savings (TTS) were common primary benefits. Several countries also captured TTS for trucks, pedestrians, and cyclists. Accident cost savings (ACC) were common, although unit accident costs varied. For secondary benefits, all countries and regions except Hong Kong and Singapore included environmental externalities. Air pollution and noise impact were common. The United States was unique in including option value, and Germany and the Netherlands were unique in adopting agglomeration benefits. The social discount rate (SDR), assessment period, and decision criteria varied. Most SDRs used the marginal rate of return on private-sector investment (yielding an SDR of 6% to 10%). Net present value was the common decision criterion, and 20 to 30 years was a common evaluation period. Standardized parameter valuations suggested commuting value of time as \$5 in Australian dollars per hour (A\$5/h) to A\$15/h, with an average of A\$10/h for public transport users and A\$10.50/h for car users. Accident cost varied; fatal accidents cost A\$0.1 to A\$4.25 million and serious accidents A\$60,000 to A\$490,000. A case study illustrates implications of adopting approaches with varied outcomes. Only Australia, the United States, the United Kingdom, and the Netherlands had positive benefit–cost ratios (1.00 to 2.61). TTS and congestion relief were major benefits (50% to 60% and 40% to 50%, respectively, of all benefits). ACC, environmental externalities, and option value benefits were not significant. Agglomeration benefits substantially increased project benefits.

Cost–benefit analysis (CBA) was developed as an assessment method for the evaluation of public policy issues and projects (1–3). Today, CBA is widely used in the evaluation of major transport investment projects, such as urban rail projects, to ensure they generate optimum returns (4, 5). The importance of CBA in the evaluation of

public transport (PT) projects is highlighted by TRB, which suggests that the increasing constraints on public funding and the sheer competition of public schemes across the whole sphere of government mean that urban rail proposals must “prove their mettle by passing strict cost–benefit assessments” (6).

According to Nash, transport was among the first fields in which CBA was regularly used as part of decision making (5). Despite this heritage significant differences remain in approaches to rail project evaluation in countries (7, 8). These differences can be of interest because

- They illustrate alternative views on CBA application,
- They can indicate new and innovative approaches to appraisals,
- They inform the development of guidelines to reflect the most appropriate methods, and
- They can illustrate points of contention within CBA application that are often a useful focus for research.

While earlier papers have compared national differences in general CBA applications (7, 8), this paper expands on this comparative analysis by

- Including more countries in the comparative framework,
- Contrasting the strategic differences and parameter valuation approaches adopted in greater detail, and
- Illustrating the implications of these differences with a case study.

This paper compares CBA guidelines for urban rail project evaluation in Australia, the United States, the United Kingdom, Canada, New Zealand, Germany, the Netherlands, France, Japan, Hong Kong, South Korea, and Singapore. A major motivation for this analysis was to understand differences in approaches adopted, as these differences might inform best practices in the field. The key findings on the different aspects of the CBA framework from a strategic viewpoint as well as the different parameter values adopted are presented. The collected evidence is largely based on published economic assessment guidelines produced by national governments to provide a general overarching framework for the appraisal of publically funded projects. For Japan, France, Germany, Hong Kong in China, South Korea, and Singapore such guidelines were not published or available; in these cases, CBA approaches were derived from Morisugi (9), Quinet (10), and Rothengatter (11) or obtained via email correspondence with the relevant authorities. Economic assessment is usually carried out as part of the local or strategic planning process as a precursor to the application of government

---

E. Gwee and G. Currie, Institute of Transport Studies, Department of Civil Engineering, Monash University, Building 60, Clayton, Victoria 3800, Australia. Alternate address for E. Gwee: Local Planning Division, Land Transport Authority, 1 Hampshire Road, Block 10, Level 2, Singapore 219428; J. Stanley, Institute of Transport and Logistics Studies, Sydney University, New South Wales 2006, Australia. Corresponding author: G. Currie, graham.currie@monash.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 73–85.  
DOI: 10.3141/2261-09

funding. Published evidence can, however, lag behind the practice of evaluation in the field. In addition, the availability of a national CBA guideline does not necessarily translate to a standardized practice within each country. For larger countries such as the United States, wide variations exist in methods, benefits categories, and parameter variation values used by localities and states. In addition, practice and guidelines change over time, so to some extent this review is a snapshot of a changing practice. These are limitations that this research has had to accept.

This paper is structured as follows:

- The second section presents a discussion of strategic differences from a rail assessment perspective in the CBA frameworks identified,
- The third section discusses parameter valuation evidence assembled in the review,
- The fourth section outlines the case study methodology adopted,
- The fifth section discusses the findings of the case study, and
- The sixth section concludes with a summary of key findings.

## RESEARCH CONTEXT

The 12 countries or regions included in this comparative assessment were selected on the basis of the following three considerations:

1. The 12 countries or regions have cities that feature highly in the Mercer *Quality of Living Survey* (12), *The Economist's World's Most Liveable Cities* (13), and Monocle's *Most Liveable Cities Index* (14). The rationale was that these renowned surveys have considered public transportation as a criterion in their ranking. It is reasonable to assume that with their good public transportation systems, these countries are likely to have also established a comprehensive assessment framework for their PT investments.
2. The selected countries included affluent Asian countries such as Singapore and South Korea and the region of Hong Kong, China, in view of their high-quality urban rail systems. The rationale was that, given their quality rail system, these countries may have also established a more comprehensive economic appraisal framework.
3. Finally, the selected countries had an availability of published national guidelines or previous research.

Table 1 outlines a comparison of CBA approaches among the 12 countries or regions studied. Key features are now discussed.

### Role of CBA

All countries generally adopt a multicriteria analysis (MCA) approach for project evaluation with CBA being one of the key components. Use of CBA varies. Project evaluation is usually supplemented with other specialized studies such as an environmental impact statement. These other factors could be considered separately from the CBA (e.g., an environmental impact assessment, as in the case of Australia) or weighted and summed up in a formal MCA (as in the case of the United Kingdom). The MCA allows decision makers to subjectively assess monetized and non-monetized impacts. Examples of the tabular summary include the Australian Appraisal Summary Table and the Japanese Benefit Incidence Table.

## Components of Financial Cost

All countries include capital, system operating, and maintenance costs in CBA. Most include costs associated with mitigation related to negative impacts of a project. In addition, the U.S. and Australian guidelines further stipulate that the costs of required improvements in other parts of the transport system that result from implementing the new rail proposal should be included as a cost component in CBA.

Only the United States, New Zealand, Germany, and Singapore do not consider the residual value of assets in CBA. The other countries treat residual value as a negative cost in the previous appraisal year to account for the benefits that the proposal can provide beyond the assessment period. Among them, the United Kingdom and the Netherlands further specify a criterion for the inclusion of residual value. Residual values are considered only for projects with finite lives less than 60 and 100 years, respectively, in the English and Dutch guidelines.

## Monetized Benefit Components

There is more variation in monetized benefit assessment in the approaches reviewed. Primary and secondary benefits are generally identified. As defined by TRB, primary benefits are directly associated with the activity of travel itself and its effect on users, while secondary benefits are generated over and above the direct benefits that accrue to users of the rail system (1, 6).

### Primary Benefits

Primary benefits comprise mainly travel time savings (TTS) and reductions in operating and accident costs. All the countries include accident costs resulting from reductions in road travel in their CBA frameworks. However, there is great variation in the unit accident costs adopted. Other benefits can be estimated for four main groups: PT users, automobile users, truck users (or cargo transit users), and pedestrians and cyclists (Table 1).

#### Public Transport Users

All countries consider TTS and fare savings as a benefit for PT users, who include existing PT users, diverted PT users (whose trips were previously made on another PT service), and former car drivers and passengers. The benefit for existing PT users is usually reflected as a savings in the generalized cost of travel, which is a function of travel time (i.e., walk, wait, transfer, and in-vehicle time) savings and fare. The benefits to diverted PT users as well as former car drivers and passengers are generally estimated as one-half of the unit benefit accruing to an existing PT user. Canada is the only country that requires a separate consideration of small travel-time savings of less than 5 min per one-way trip. They view that a small TTS is unlikely to be put to any productive use. Hence, while such benefits are clearly identified, they are not included in the CBA calculation.

#### Automobile Users

Congestion relief is estimated in CBA by all the countries except Japan. This benefit is measured largely in terms of TTS as well as a

TABLE 1 Comparison Matrix of International Assessment of Evaluation Approaches to Urban Rail Projects (1; 6–11; 15–31)

		Country or Region											
Parameter	Details	Australia	New Zealand	United States	Canada	United Kingdom	France	Netherlands	Germany	Japan	South Korea	Hong Kong	Singapore
Cost	Capital costs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Operating and maintenance costs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Costs to mitigate negative effects of project	✓	✓		✓	✓		✓	✓		✓	✓	
	Costs to improve other parts of transport system	✓		✓									✓
	Residual value of asset	✓			✓	✓	✓	✓		✓	✓	✓	
Benefit	PT users												
	Travel time savings	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Fare savings and out-of-pocket savings	✓	✓	✓	✓	✓	✓	✓	✓	✓			
	Auto users												
	Travel time savings	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
	Operating cost savings	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
	Vehicle ownership and maintenance cost savings	✓		✓									
	Truck users												
	Travel time savings	✓	✓	✓	✓						✓	✓	✓
	Operating cost savings	✓	✓	✓	✓						✓	✓	✓
	Pedestrians and cyclists												
	Travel time savings	✓		✓									
	Vehicle ownership and maintenance cost savings	✓		✓									
	Externalities												
	Air pollution	✓	✓	✓	?		✓	✓	✓	✓	✓		
	Greenhouse gas emission	✓	✓		?		✓	✓		✓	✓		
	Water quality impact	✓		✓	?			?					
	Noise impact	✓	✓	✓	?	✓	✓	✓	✓	✓	✓		
	Impact to nature and landscape	✓			?			?					
	Others												
	Accident cost savings	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Savings in avoided costs	✓	✓	✓					✓				
Option value			✓										
Agglomeration benefits		✓				✓		✓					
Enhancement to property values							?						

(continued on next page)

TABLE 1 (continued) Comparison Matrix of International Assessment of Evaluation Approaches to Urban Rail Projects (1; 6–11; 15–31)

Parameter	Details	Country or Region											
		Australia	New Zealand	United States	Canada	United Kingdom	France	Netherlands	Germany	Japan	South Korea	Hong Kong	Singapore
Discount rate	Method of derivation												
	Return on private investment		✓	✓	✓					X	✓	✓	
	Time preference					✓				X			✓
	Borrowing rate	✓						✓		X			
	Weighted average approach									X			
	Shadow price of capital approach						✓			X			
	Optimal growth rate approach								✓	X			
	Value adopted	Discount rate nominated by funding body	8%	3% and 7%	10%	2.5%–3.50%	8%	4%	3%	4%	6.5%	Depend on rail operator	4%
Evaluation period (years)		50	30	20	30	60	30	Tied to project lifespan (max. 100)	40	30–50	20–30	30–120	60
Decision criteria	NPV	✓		✓	✓	✓	✓	✓			✓	✓	
	BCR	✓	✓						✓	✓	✓		✓
	IRR						✓	✓			✓	✓	

NOTE: ✓ = approach adopted in guidelines from country; blank = monetized and included; ? = no information or unclear whether subject parameter is included in CBA; X = not monetized; NPV = net present value; BCR = benefit–cost ratio; IRR = internal rate of return.

SOURCE: K. S. Kim, personal communication, 2008; W. L. Lim, personal communication, 2008; S. Zwartjes, personal communication, 2007; A. I. J. M. van der Hoorn, personal communication, 2008.

savings in automobile operating costs for those drivers who chose to continue to stay on the road network after the implementation of the rail initiative. Australia and the United States take the view that transit improvements may affect relatively long-run decisions, including the decision to own motor vehicles. Hence savings for vehicle ownership and the subsequent maintenance costs are considered in their CBA approaches.

#### *Truck Users and Freight or Cargo Transit*

The United States, New Zealand, Canada, Australia, Hong Kong, and Singapore include TTS and operating cost reductions as well as savings in time-related inventory costs for truck users (or goods vehicles) in CBA. Clearly this parameter is a valid factor when road freight volumes are significant or when the cargoes involved are valuable (1). These benefits are estimated explicitly in the U.S., New Zealand, and Canadian guidelines. In the Australian, Hong Kong, and Singapore approaches, freight impacts are estimated from the average resource value of time for goods vehicles, which is an input in the estimation of the TTS for these vehicles.

#### *Pedestrians and Cyclists*

Explicit consideration of the impacts on former cyclists and pedestrians who switch to PT are included in the Australian and American CBAs. However, the guidelines involved caution these impacts should only be considered if they are expected to be substantial. Other countries such as the United Kingdom assesses such impacts qualitatively and do not include them explicitly within CBA.

### **Secondary Benefits**

Secondary benefits are generated over and above the direct benefits that accrue to users of the rail system and largely comprise the positive effect to environmental externalities such as air quality and noise impact (1). In addition, several countries also include wider economic benefits such as option value (OV) and agglomeration benefits into CBA.

#### *Environmental Externalities*

All countries except Hong Kong and Singapore have evaluated and monetized environmental externalities in CBA. These include air pollution and noise impacts. Approaches in Canada and the Netherlands stress the need to quantify environmental impacts although no specific methodology to do this is detailed in their guidelines. The New Zealand guidelines have the longest checklist of environmental impacts (including visual impact as well as overshadowing impact) to be examined as part of their project evaluation. However, these impacts are examined separately from the overall quantitative CBA appraisal. The U.K. guidelines also recommend the assessment of environmental impacts outside the CBA framework.

#### *Impact of Air Pollution and Greenhouse Gas Emissions*

Most countries consider the intensity of pollutants including carbon monoxide, particulate matter smaller than 10 µm in diameter, hydro-

carbons, and oxides of sulfur and nitrogen. Japan measures only the level of nitrogen oxide emissions. Air pollution and greenhouse gas impacts are included in CBA approaches in Australia, New Zealand, France, and Japan.

#### *Impact of Noise*

Internationally, noise impact is one of the two most common externalities included in CBA. Most countries estimate a monetary value by using a hedonic pricing method approach except Germany, which bases its value on the cost for equipping houses with noise-proof glazing.

#### *Impact of Water Quality and Nature*

Water quality and nature impacts are included in CBA approaches in selected countries. Water quality impact is addressed in the U.S. and Australian guidelines. Canada and the Netherlands recommend measuring these impacts as far as possible, while in New Zealand this impact is assessed in a study separate to the CBA. There is no mention of these impacts in the CBA guidelines of the other countries assessed.

Only Australia requires the quantification of nature impacts in CBA. New Zealand guidelines assess nature impacts separately from CBA. Other countries include the cost of the mitigating impacts on nature measures as part of the project costs. Hence, while these countries do not monetize the impact of transport on nature directly as benefits, these impacts are assessed indirectly in terms of mitigation costs and incorporated in the overall project costs.

#### *Option Value*

OV is the “willingness-to-pay to preserve the option of using a transport service for trips not yet anticipated or currently undertaken by other modes over and above the expected value of any such use” (32). Only the United States includes OV in its appraisal framework.

#### *Agglomeration Benefits*

The other area of recent international interest is the consideration of wider agglomeration benefits in transport appraisals (4). Agglomeration benefits result from the increase in productivity, creativity, and synergy among firms caused by a higher concentration of firms or higher density of employment made possible by more compact, transit-served development (6, 33). Only the guidelines of the United Kingdom, New Zealand, Germany, and the Netherlands have incorporated this benefit in their CBA to various extents (Table 1).

According to Keegan et al. (4), no significant research has been performed in this area until the recent work of Graham (34–38), who suggests that the transport system can influence the proximity between firms and the labor market to a certain extent and can improve urban or industrial densities by rendering a larger scale of activity more accessible. A major transport study that included agglomeration effects was the recent U.K. study by Sir Rod Eddington (4, 39). Eddington noted that the agglomeration effect could increase the overall project benefits of transport projects “by up to 50 per cent in some cases” (39). Vickerman agreed and estimated that the wider economic benefits “generated by rail projects may

amount to as much as 55 per cent of the direct transport benefits” (33). Despite these findings, agglomeration benefits are not included in U.K. guidelines or in CBA approaches for most countries.

### *Enhancement to Property Values*

The explicit enhancement of property values as part of CBA is suggested in the Netherlands. A. I. J. M. van der Hoorn (personal communication, 2008) and S. Zwartjes (personal communication, 2007) recommend the inclusion of enhancement to property values for larger Dutch transport projects, such as the high-speed rail link from Schiphol to Groningen. However, these benefits are not mentioned in the Dutch national guidelines. Other guidelines have suggested this is double counting of travel time (user) benefits.

## **Accounting Approaches**

### *Social Discount Rate*

Table 1 illustrates a range of approaches to social discount rate (SDR) development. These are real, not nominal rates, that is, they are adjusted for inflation. The most common approach is the marginal rate of return on private-sector investment, which is adopted by the United States, New Zealand, Canada, Hong Kong, and South Korea. This approach yields an SDR of 6% to 10%. Only the United Kingdom stipulates a different SDR for different evaluation periods:

- 3.5% (for 0 to 30 years),
- 3.0% (for 31 to 75 years), and
- 2.5% (after 75 years), although transport projects are unlikely to require appraisal that far into the future.

The Australian guidelines do not specify an SDR but recommend that the evaluation use the SDR nominated by the funding body. Hong Kong uses a similar approach. Only the Canadian guidelines state an explicit range of discount rates for sensitivity analysis.

### *Assessment Period*

The most common evaluation period appears to be 20 to 30 years. The Netherlands and Hong Kong tagged their assessment period to the types of transport projects being evaluated. In these cases evaluations up to 100 years or more can be carried out. The wide variation in the analysis period is interesting and is a potential area of future research.

### *Decision Criteria*

Three decision rules are most commonly employed in CBA:

- Net present value (NPV),
- Benefit–cost ratio (BCR, also known as the profitability index), and
- Internal rate of return (IRR).

NPV is the most common decision criterion. IRR is always taken as a second decision criterion by countries apart from the NPV. No

countries adopt the IRR as the only decision criterion. France and New Zealand use the first-year rate of return, which is equal to the surplus of the first year divided by the cost of the investment, to check the year of implementation of their projects. The optimum year of implementation is when the first-year rate of return is equal to the discount rate.

Most countries require a BCR greater than one. In Germany a BCR value greater than three is adopted. Rothengatter (11) points out that this is a result of expected double counting of effects in the German method. He suggested that double counting of effects was deliberately built into the German system to mitigate equity issues among regions.

## **PARAMETER VALUATIONS**

This section examines parameters included in rail CBA approaches in three broad categories: value of time (VOT), accident costs, and values of externalities (Table 2).

To aid comparison, parameter values for each country are converted to 2006 Australian dollars. As willingness to pay (WTP) is a unifying element in values, the change in the average wage of each country provides a good approximation of how WTP would vary over time. The parameter values were first updated to 2006 values according to the average wage increment of each country between the date the VOT was captured (as shown in the guidelines) and 2006. The 2006 values were then converted to Australian currency according to the exchange rates provided by the Australian Reserve Bank.

### **Value of Time**

VOT data are for PT users and car drivers for trips to and from work. This time bracket is chosen as it represents the time when passenger ridership and vehicular traffic are the highest in the network and VOT is expected to have the most significant implication to the appraisal. Two methods for VOT are adopted in the CBA approaches examined:

- The wage rate approach and
- The stated preference or revealed preference approach.

VOT for commuting trips to and from work ranges from about \$5 in Australian dollars per hour (A\$5/h) (for Singapore and Hong Kong) to A\$15/h (for the Netherlands) (Figure 1). This significant discrepancy in VOT can largely be attributed to the different wage rate premiums of each country. The implication is that countries like the Netherlands would value benefits three times more than Singapore for the same project TTS. This implication is significant given the dominant role of TTS in urban rail appraisals.

Most of the countries adopt a similar VOT for PT users and car users. In the United States, for example, VOT for both PT and car users is A\$14.01 [based on the average U.S. wage of \$10/h (1)]. For New Zealand, Hong Kong, and Singapore, a higher VOT is used for car users than for PT users, while the Netherlands adopts a slightly lower VOT for car users. The use of different VOTs for automobile travel and transit travel is interesting and suggests future research. The observed average VOT for commuting to and from work of these countries is about A\$10 for PT users and A\$10.50 for car users. VOT of most countries is about 30% to 50% of the average hourly wage rate.

## Accident Costs

Two approaches are used to estimate accident costs in the CBAs examined:

- The human capital approach and
- The WTP approach.

The human capital approach involves estimating the discounted present value of all costs arising from a crash that can be directly measured, including the loss of future earnings. The WTP approach involves estimating the monetary amount that people are willing to forgo to reduce the risk of death or injury (16).

Valuation of mortality accidents is very different among countries (Table 2 and Figure 2). The highest value of a fatal accident is A\$6.66 million in the United States; the lowest is A\$0.1 million in Singapore. Valuation of serious accidents and minor accidents also varies widely among countries. While each serious accident costs about A\$490,000 in the United Kingdom, the value is about eight times lower in Germany at about A\$60,000. Likewise, minor accidents are about nine times more costly in the United Kingdom and Canada compared with Germany and Japan. The U.S. approach applies the same value to serious accidents and minor accidents, both of which were valued at \$10,000 (2002 values) (1).

## Externality Costs

Broad average valuations for externality costs are shown in Table 2. Most guidelines suggest that these values should be used with caution and that a detailed assessment should be conducted if certain externalities are expected to have a significant impact on the appraisal.

Nash notes that one of the biggest challenges in deriving a monetary value for environmental impacts is "that different studies tend to come up with totally different results" as a result of the different methods employed as well as the differing principles in the way costs are assessed (40). This challenge is clearly evident from Table 2, which shows that different countries have derived very different monetary values (or a range of values) for environmental impacts. For example, the value of noise pollution is estimated to be about A\$0.0015/vehicle km in the United States, but the Australian guidelines estimate a value of A\$0.0080/vehicle kilometer, which is about five times higher. Australia, New Zealand, and the Netherlands have a unified cost value for air pollution, but the United States uses a different monetary value for each air pollutant.

## CASE STUDY APPROACH

To illustrate the implications of the different CBA approaches examined an example rail improvement project is evaluated. The case study undertaken is a rail electrification project in Melbourne, Australia. This project involved the electrification of about 30 km of existing rail tracks as well as the upgrading of the corresponding facilities and vehicles at an estimated project cost of about A\$80 million. The project is expected to benefit some 2,000 rail users and 2,500 automobile users daily. All CBA approaches are applied to the case study excluding those from the South Korea, where no parameter value information was available. In summary, the case study was prepared using the following steps:

- CBA for each country is applied according to the principles defined in Table 1 and the parameter values shown in Table 2.

- To aid comparison, all appraisals are carried out in equivalent 2006 Australian dollars. The costs of each country were converted to the 2006 value of their respective currencies by using wage trends from the relevant national authorities. These were then converted to Australian dollars by using the 2006 exchange rates from the Australian Reserve Bank.

- Capital and operating costs of the hypothetical rail project are assumed to be the same for all 12 countries. (While countries with lower VOT will also have lower labor costs, there is insufficient data for this study to work out how the project costs may vary among countries. This research therefore made the above assumption to facilitate the study. This estimate is envisaged to yield a conservative result since the project costs may be on the high side for countries with lower labor costs.)

- Vehicle operating cost (VOC) values in the guidelines of Australia, New Zealand, the United States, the United Kingdom, Canada, and the Netherlands range between A\$0.14/km and A\$0.18/km for passenger cars commuting to and from work. An average VOC of A\$0.16/km is adopted for countries lacking published VOC values.

- For Canada and the Netherlands, which have indicated that environmental impacts should be quantified and valued as far as possible, a reduction in air pollution, greenhouse effects, and noise pollution are included in the economic appraisal.

- For the United States, which includes OV's in its CBA, an OV benefit is estimated on the basis of information provided by TRB (1).

- For the United Kingdom, New Zealand, the Netherlands, and Germany, which incorporate agglomeration effects in their CBAs, an agglomeration benefit is benchmarked at 50% of the direct transport benefits as suggested by Eddington (39) and Vickerman (33).

- An SDR of 6% and 10% is assumed for the Australian and Hong Kong appraisals, respectively. This assumption is based on the understanding that the Australian rate is tagged to the government's borrowing rate, while that of Hong Kong is based on the marginal rate of return of private investment.

- For those countries for which parameter values were not available, the corresponding Australian values were adopted.

## RESULTS AND DISCUSSION

Table 3 shows the resulting CBA assessments obtained by using the above approach. Very different evaluation outcomes emerge. Only the Australian, U.S., U.K., and Dutch approaches found the proposal economically feasible. Although Germany achieved a BCR of 1.31, the proposal was not economically feasible as their decision criteria require a project to have a minimum BCR of three before it will be considered for implementation.

BCR of the countries that assessed the project as feasible ranged from 1.00 (Australia) to 2.61 (the Netherlands). This significant difference is largely because of the higher VOT adopted in the Netherlands as well as the inclusion of agglomeration benefits in its CBA. The following discussion considers various CBA components and the results associated with these components.

### Value of Time

TTS contributes about 50% to 60% of the total benefits generated by the project. On this basis appraisal outcomes can be very sensitive

TABLE 2 Comparison of Parameter Values for 2002 (1; 6–11; 15–31)

Parameter	Country or Region					
	Australia	New Zealand	United States	Canada	United Kingdom	France
VOT Commuting to and from Work						
PT users (A\$/h)	10	6.42	14.01	9.56	12.10	11.41
Car users (A\$/h)	10	7.59	14.01	9.56	12.10	411.41
Accident Costs						
Fatal accident (A\$ millions)	1.85	2.99	6.66	2.39	4.25	1.24
Serious accident (A\$)	454,230	320,970	382,950	132,827	487,535	127,840
Slight accident (A\$)	14,014	18,723	13,320	50,313	49,677	27,179
Externalities						
Air pollution	A\$0.0252/vkm	A\$0.0089/vkm	A\$0.02/kg CO to A\$20.40/kg PM10	X	Θ	A\$0.0268/pax km
Greenhouse gas emission	A\$0.0031/vkm	7.5% of VOC	Θ			A\$0.0081/pax km
Water quality impact	A\$0.0038/vkm	Θ	A\$0.0005/vkm			Θ
Noise impact	A\$0.0080/vkm	A\$365.55/db/household/year	A\$0.0015/vkm		A\$24.38 to A\$284.40/db/household/year	X
Nature and landscape	A\$0.0034/vkm	Θ	Θ		Θ	Θ

NOTE: X = no information; vkm = vehicle kilometer; Θ = not assessed; CO = carbon monoxide; PM = particulate matter; db = decibel; pax km = passenger kilometer.

SOURCE: K. S. Kim, personal communication, 2008; W. L. Lim, personal communication, 2008; S. Zwartjes, personal communication, 2007; A. I. J. M. van der Hoorn, personal communication, 2008.

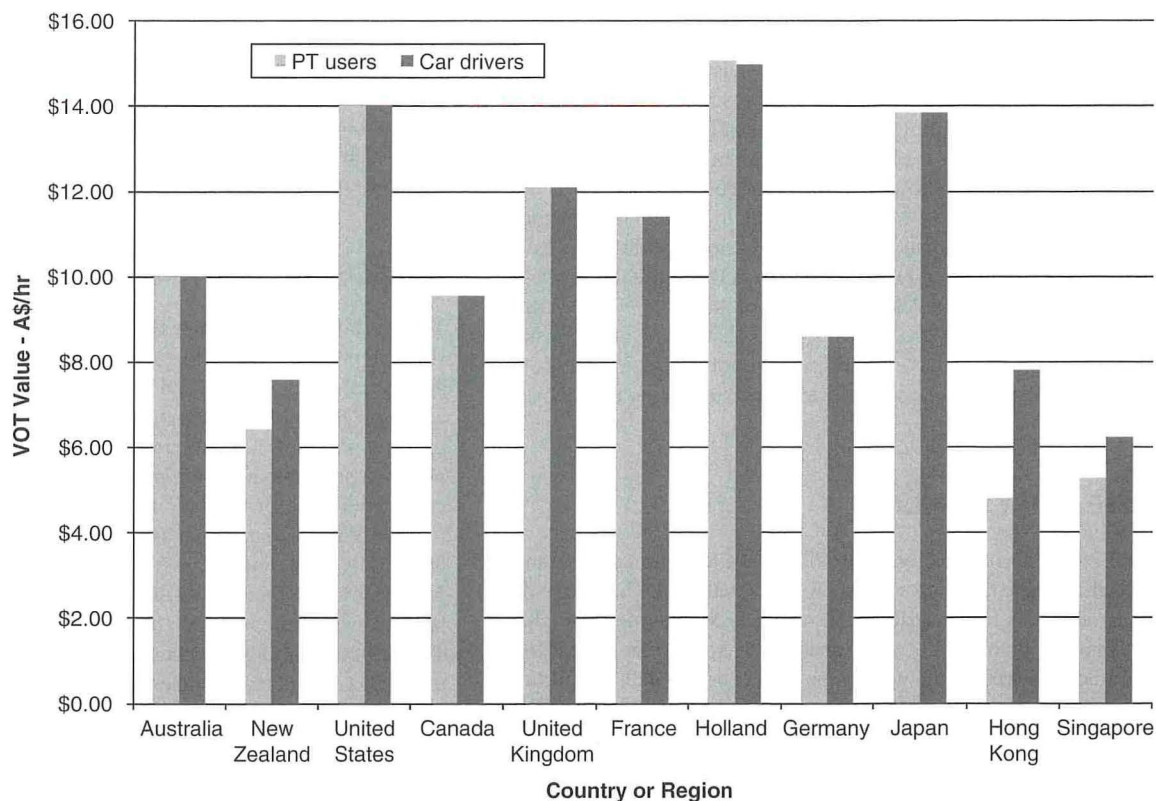


FIGURE 1 Standardized value of time results for car and PT users.

Netherlands	Germany	Japan	South Korea	Hong Kong	Singapore
15.07	8.59	13.85	X	4.81	5.27
14.97	8.59	13.85		7.81	6.26
1.55	1.33	0.32	X	X	70,000–100,000
206,464	60,663	96,866			
30,970	5,257	6,183			
A\$0.0134/vkm	X	A\$5,977/ton nitrogen oxide	X	Θ	Θ
A\$0.0103/vkm	Θ	A\$24/ton carbon	X		
X	Θ	Θ	Θ		
A\$0.0052/vkm	X	A\$4,897/db/km/year	X		
X	Θ	Θ	Θ		

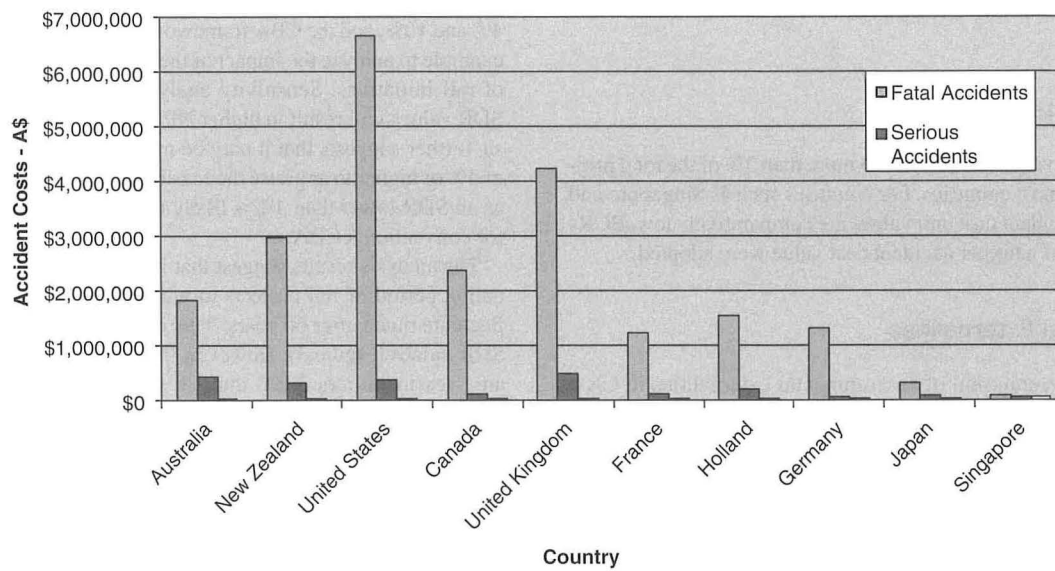


FIGURE 2 Standardized value of accident costs.

TABLE 3 Evaluation Results of Case Study

Detail	Country or Region					
	Australia	New Zealand	United States	Canada	United Kingdom	France
SDR used in case study (%)	(assumed) 4	10	3 and 7	10	2.5–3.5	8
Evaluation period (years)	50	25	20	30	60	30
Evaluation results						
NPV (A\$ millions)	112	–27,233	5,927	–40,583	130,836	–9,912
BCR	1.00	0.69	1.07	0.51	2.09	0.90
IRR (%)						5.50
Evaluation outcome	✓	—	✓	—	✓	—

NOTE: ✓ = passed economic appraisal; — = criteria not defined.

to changes in VOT. For example, if the VOT used in the U.S. assessment is increased by merely \$0.10/h, the project would be economically feasible with a BCR above unity. Clearly it is important that VOT estimates are updated periodically to ensure that benefits are accurately assessed against costs.

### Relief for Congestion

Most countries included road user TTS and vehicle operating costs when estimating congestion relief (16, 22). Congestion relief accounted for about 40% to 50% of total project benefits. Japan does not include congestion relief benefits. This is a major omission since its BCR would improve significantly from about 0.48 to 0.55 to about 1.21 to 1.41 if congestion relief were included in its assessment.

### Accident Costs

Accident cost savings contributed no more than 3% of the total project benefits for most countries. For countries such as Singapore and Japan, where accident cost unit values are comparatively low, BCRs would improve if a higher accident cost value were adopted.

### Environmental Externalities

The inclusion or omission of environmental externalities in CBA was not a significant factor in the outcome of the rail case study, an observation consistent with the views of TRB (1). Environmental benefits in the case study accounted for about 4% or less of the total project benefits.

### Option Value

OV benefits in the U.S. CBA were estimated for scenarios in which car users were willing to buy the options to use the rail alternative two to 10 times a year according to estimates from TRB (1). For all scenarios, OV benefits accounted for no more than 1% of total project benefits on this basis. However, it is probably premature to conclude from this that OV benefits are not important. Laird et al. note that “the field of measuring transport option values clearly is far from developed” (32), and further research in this area is warranted.

### Agglomeration Benefits

Agglomeration benefits are included in the Dutch and German CBA analyses. Results suggest that these benefits could account for as much as 30% of total project benefits. If they were included in the CBA approaches for all countries, the appraisal outcomes would improve significantly. Given the lack of research in this area (4) and the potential significance of its contribution to the economic viability of transport proposals, further research into the validity of agglomeration benefits is indeed important.

### Social Discount Rate and Assessment Period

Sensitivity analysis was performed by using discount rates between 4% and 10%, and the CBA framework of Hong Kong was used as an example to analyze the impact of the assessment period on the CBA of rail initiatives. Sensitivity analysis results indicate that lower SDR values will result in higher NPV values (Figure 1). The analysis further suggests that it may be more appropriate to use an SDR of 4% or higher to appraise the feasibility of individual rail projects, as an SDR lower than 4% is likely to always yield a positive NPV for conventional CBA.

The analysis results suggest that it is preferable to keep the evaluation period of rail projects to within 60 years as NPV does not fluctuate much after 60 years. This is especially the case for higher SDR values (Figure 3). However, the case study also suggests that an assessment period less than 30 years may be too short for rail investment appraisal as the benefits accrued over 30 years do not act to cover the substantial rail project costs incurred earlier in the project's life.

### CONCLUSIONS

This paper compares CBA approaches to urban rail project evaluation in Australia, the United States, the United Kingdom, Canada, New Zealand, Germany, the Netherlands, France, Japan, Hong Kong, South Korea, and Singapore. An assessment of strategic frameworks and individual parameter valuations was undertaken. A case study evaluation illustrates the impacts of differences in CBA approaches.

Strategically, all countries adopt an MCA for project evaluation, with CBA being one of the key components. There is some variation in how CBA is used. In terms of cost components, all the countries

Netherlands	Germany	Japan	South Korea	Hong Kong	Singapore
4	3	4	6.5	(assumed) 10	4
(assumed) 50	40	30–50	20–30	30–120	60
180,850		–53,987	Not evaluated	–60,155	–65,771
2.61	1.31	0.48–0.55		0.27	0.43
15.50				5	
✓	—	—	—	—	—

consider capital, operating, and maintenance costs in their CBAs. Some also include the residual value of the assets in their assessment.

There is noticeably more variation in approaches to monetized benefits. TTS is a common primary benefit. Japan does not include congestion relief in its CBA. Several countries, including Australia, also capture TTS for truck users as well as pedestrians and cyclists in their assessments. Another common benefit included in CBA internationally is accident cost savings. However, there is much variation in the unit accident costs adopted. The use of sensitivity analysis of parameter values in the evaluation would enable a more robust set of analyses when input values vary between countries.

In terms of secondary benefits, all countries and regions except Hong Kong and Singapore include the impact to environmental externalities to various extents in their CBAs. Air pollution and noise impact are the most common environmental externalities considered. The United States is the only country that includes OV in its CBA, and only Germany and the Netherlands adopt agglomeration benefits in their economic appraisals.

There are considerable variations in SDR, the assessment period, and the decision criteria adopted. Most countries derive SDR on the

basis of the marginal rate of return on private-sector investment, which yields an SDR of 6% to 10%. The most common evaluation period is 20 to 30 years. NPV is the most common decision criterion among the approaches examined.

Parameter valuations were assessed in comparative terms by standardizing to a single currency and year of estimate. VOT for commuting trips to and from work range from A\$5/h to A\$15/h, with an average value of about A\$10/h for PT users and A\$10.50/h for car users. Likewise, the accident cost unit values vary between countries. The value of a fatal accident ranges from A\$0.1 to A\$4.25 million; serious accidents vary between A\$60,000 to A\$490,000. There is also much variation in the monetary valuations for the unit environmental impacts.

To illustrate the implications of these strategic and tactical differences, a case study was developed that found very different evaluation outcomes according to the approaches adopted. Only the approaches in Australia, the United States, the United Kingdom, and the Netherlands found the proposal economically feasible (BCR between 1.00 and 2.61). The most important benefits identified from this analysis were TTS and congestion relief, which contributed about 50% to 60%

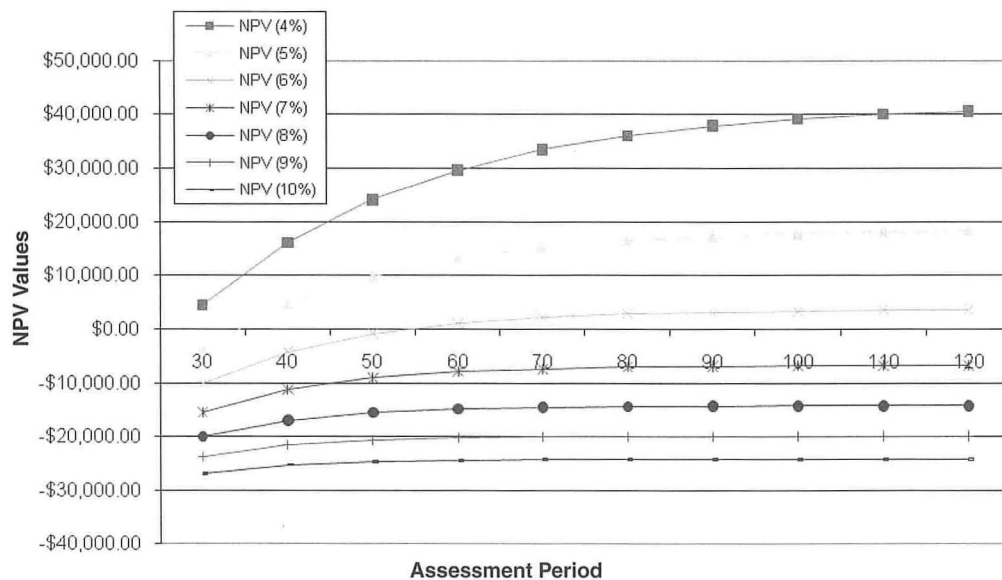


FIGURE 3 NPV trends for Hong Kong across different SDRs and assessment periods.

and 40% to 50% of total project benefits, respectively. Accident cost savings contributed no more than 3% of total project benefits, and the results of the case study suggest that the impact of environmental externalities in CBA is not significant.

The case study analysis suggests that OV benefits accounted for no more than 1% of total project benefits in the U.S. appraisal. However, it is acknowledged that OV is a relatively new area in transport economics, and it is probably too early to conclude that OV benefits are not important. The case study also found that agglomeration benefits would substantially increase project benefits if they were included in CBA approaches.

Sensitivity analysis suggests that an SDR of 4% or higher should be used to assess the economic feasibility of individual rail proposals, and 30 to 60 years appears to be the most common assessment period for rail investment projects.

So what do the findings say about international approaches to CBA in urban rail? The diversity of approaches clearly reflects differing views of what methods are appropriate in different geographies. This might be because of specific conditions in some locations; for example, OVs may not be considered important in one location but important in others. Another example is the diversity of accident costs: it may well be that fatality cost assessments vary because of different views on the value of life in different locations. However, some of the findings seem hard to interpret from this perspective. Neither Hong Kong nor Singapore includes environmental externalities, and Japan does not include congestion relief benefits. It seems unlikely that such factors are unimportant in high-density cities with urban pollution and endemic traffic congestion. The use of rail in each of these Asian locations is likely to be fundamental to addressing environmental and congestion relief issues, yet authorities in these contexts do not consider these issues important to rail appraisal.

Another possible explanation for diversity in approaches is the lack of uptake or understanding of new approaches to CBA appraisal methods. Conservatism in approaches to evaluation, especially a preference for historical and conventional methods to new, untried, and possibly speculative methods, is another explanation. This can be a common feature of national treasuries seeking balanced approaches to fiscal management.

The variation in CBA approaches can suggest possible improvements, particularly in cases in which countries have chosen to adopt different kinds of benefits. The obvious examples are those countries not using certain kinds of benefits, such as environmental and congestion relief. Unusual parameter values can also point to the need for changes in approach. When values adopted are unusual compared with international practice there is a need to understand why this is the case. These are all promising outcomes from the comparative assessments performed in this study.

To some extent CBA approaches do not have to conform to any international standards when being applied to project appraisal for local funds. All local projects being compared have the same local approach, and results are technically comparable. Only when methods are compared between countries do differences exist. While this is undoubtedly true, capital for many major rail projects is often sourced from international financial institutions. Moreover, international companies are often involved in construction and operations, so there is at least some rationale for comparable CBA approaches on an international basis.

When rail projects are compared with road projects and projects from other sectors, both what is included and excluded in the evaluation may be significant to the final CBA. Agglomeration benefits, for example, are important to rail in the Netherlands and would con-

siderably boost the net benefits of rail projects there. Countries or regions that don't include these benefits run the risk of undervaluing rail projects. Similarly, rail projects evaluated in Hong Kong and Singapore without consideration of environmental benefits or Japanese rail projects, which omit congestion relief benefits, may suffer an adverse CBA.

The variation in approaches is also an important consideration for the profession and discipline of economics. It seems unlikely that the methodological diversity demonstrated in this paper can all be based on a logical assessment of differences in local conditions. There seems to be a role for economists to encourage more standardization of approaches among countries. Methods that adjust for variation in values adopted (such as sensitivity analysis) should be encouraged to ensure a more robust analysis. Disagreement about new areas of economic benefits such as agglomeration economies strongly suggest the need for research to clarify the viability of these considerations and to develop methods that are both robust and easy to apply in measuring benefits of this kind. There is also a role for research to explore the rationale behind the adoption of CBA approaches in different countries. In this way variations in approaches to CBA in urban rail might rest on a defensible knowledge base that might also identify areas in which additional research in CBA methods are required.

## REFERENCES

1. *TCRP Report 78: Estimating the Benefits and Costs of Public Transit Projects: A Guidebook for Practitioners*. TRB, National Research Council, Washington, D.C., 2002.
2. Nas, T. *Cost-Benefit Analysis: Theory and Application*. Sage Publications, Inc., Thousand Oaks, Calif., 1996.
3. Boardman, A. *Cost-Benefit Analysis: Concepts and Practice*, 3rd ed. Prentice-Hall, Upper Saddle River, N.J., 2006.
4. Keegan, M., G. Favero, and C. Porter. Transport Appraisal: Are We Including All the Benefits? *UITP Magazine PT International*, November–December 2007.
5. Nash, C. Cost-Benefit Analysis of Transport Projects. In *Efficiency in the Public Sector: The Theory and Practice of Cost-Benefit Analysis* (A. Williams and E. Giardina, eds.), Edward Elgar, Aldershot, Hampshire, United Kingdom, and Brookfield, Vt., 1993.
6. *TCRP Report 35: Economic Impact Analysis of Transit Investments: Guidebook for Practitioners*. TRB, National Research Council, Washington, D.C., 1998.
7. Hayashi, Y., and H. Morisugi. International Comparison of Background Concept and Methodology of Transportation Project Appraisal. *Transport Policy*, Vol. 7, No. 1, 2000, pp. 73–88.
8. Nakamura, H. The Economic Evaluation of Transport Infrastructure: Need for International Comparison. *Transport Policy*, Vol. 7, No. 1, 2000, pp. 3–6.
9. Morisugi, H. Evaluation Methodologies of Transportation Projects in Japan. *Transport Policy*, Vol. 7, No. 1, 2000, pp. 35–40.
10. Quinet, E. Evaluation Methodologies of Transportation Projects in France. *Transport Policy*, Vol. 7, No. 1, 2000, pp. 27–34.
11. Rothengatter, W. Evaluation of Infrastructure Investments in Germany. *Transport Policy*, Vol. 7, No. 1, 2000, pp. 17–25.
12. Mercer Consulting Group. *Quality of Living Survey 2009*. July 22, 2009. <http://www.mercer.com>.
13. The Economist Intelligence Unit. *The Economist's World's Most Liveable Cities 2009*. July 22, 2009. <http://www.economist.com/markets/rankings>.
14. Monocle. *Most Liveable Cities Index 2009*. July 22, 2009. <http://www.monocle.com>.
15. Lam, A. *The Cost-Benefit Analysis Framework of Hong Kong* (E. Gwee, ed.). Melbourne, Australia, 2007.
16. *National Guidelines for Transport System Management in Australia*. Australian Transport Council. Canberra, Australia, 2006.
17. Eijgenraam, C. J. J. *Evaluation of Infrastructural Projects: Section 1: Main Report*. Ministry of Transport Public Works and Water Management and Ministry of Economic Affairs, The Hague, Netherlands, 2000.

18. Eijgenraam, C. J. J. *Evaluation of Infrastructural Projects: Section 2: Capita Selecta*. Ministry of Transport Public Works and Water Management and Ministry of Economic Affairs, The Hague, Netherlands, 2000.
19. *Procedures and Technical Methods for Transit Project Planning (Alternatives Analysis Technical Guidance)*. FTA, U.S. Department of Transportation, 2007.
20. Her Majesty's Treasury. *The Green Book: Appraisal and Evaluation in Central Government*. Treasury Guidance, London, 2003.
21. *Economic Evaluation Manual, Volume 2*. Land Transport New Zealand, Wellington, New Zealand, 2005.
22. *Economic Evaluation Manual, Volume 1*. Land Transport New Zealand, Wellington, New Zealand, 2007.
23. Lee, D. B., Jr. Methods for Evaluation of Transportation Projects in the USA. *Transport Policy*, Vol. 7, No. 1, 2000, pp. 41–50.
24. *Guidelines and Discount Rates for Benefit–Cost Analysis of Federal Programs*. Office of Management and Budget, Washington, D.C., 1992.
25. *Guideline on Economic, Social and Environmental Cost–Benefit Analysis*. Planning and Policy Division, Department of Infrastructure, Melbourne, Australia, 2005.
26. *Guide to Benefit–Cost Analysis in Transport Canada*. Transport Canada, Ottawa, Ontario, Canada, 1994.
27. *TCRP Report 20: Measuring and Valuing Transit Benefits and Disbenefits*. TRB, National Research Council, Washington, D.C., 1996.
28. *Benefit–Cost Analysis Guide*. Treasury Board of Canada Secretariat, Ottawa, Ontario, Canada, 1998.
29. Vickerman, R. Evaluation Methodologies for Transport Projects in the United Kingdom. *Transport Policy*, Vol. 7, No. 1, 2000, pp. 7–16.
30. Young, L. Determining the Discount Rate for Government Projects. Working Paper 02/21. New Zealand Treasury, Wellington, New Zealand, 2002.
31. *Update of RUC Unit Values to June 2005*. Austroads Technical Report AP-T70/06. Austroads, Sydney, Australia, 2006.
32. Laird, J., K. Geurs, and C. Nash. *Option and Non-Use Values and Rail Project Appraisal*. Department for Transport, London, 2007.
33. Vickerman, R. Provision of Public Transport Under Conflicting Regulatory Regimes. In *10th International Conference on Competition and Ownership in Land Passenger Transport*, Hamilton Island, Australia, 2007.
34. Graham, D. *Wider Economic Benefits of Transport Improvements: Link Between Agglomeration and Productivity*. Stage 1 report. Center for Transport Studies, Imperial College London, London, 2005.
35. Graham, D. *Wider Economic Benefits of Transport Improvements: Link Between Agglomeration and Productivity*. Stage 2 Report. Center for Transport Studies, Imperial College London, London, 2006.
36. Graham, D. *Investigating the Link Between Productivity and Agglomeration for UK Industries*. Center for Transport Studies, Imperial College London, London, 2005.
37. Graham, D. Agglomeration, Productivity and Transport Investment. *Journal of Transport Economics and Policy*, Vol. 41, No. 3, 2007, pp. 317–343.
38. Graham, D. An Empirical Analytical Framework for Agglomeration Economies. *Annals of Regional Science*, Vol. 42, No. 2, 2008, pp. 267–289.
39. Eddington, R. *The Eddington Transport Study—The Case for Action: Sir Rod Eddington's Advice to Government*. Her Majesty's Treasury, London, 2006.
40. Nash, C. Transport Externalities: Does Monetary Valuation Make Sense? In *Recent Developments in Transport Economics* (G. D. Rus and C. Nash, eds.), Ashgate Publishing Limited, Aldershot, Hampshire, United Kingdom, 1997.

---

*The Transportation Economics Committee peer-reviewed this paper.*

# Influence of Value of Time on Profitability of Railway Projects

## Theoretical Formulation and Case Study

O. Criado, M. Sánchez-Borràs, and F. Robusté

The recent worldwide financial crisis has imposed limitations on public and private spending on new transport infrastructures. As a result, there is a need for new stricter criteria to prioritize projects in a more socioeconomically efficient way. Project appraisal is a common tool used for this effort. The value of time (VOT) is a relevant element in the appraisal of transport projects. VOT is essential for the calculation of the gain derived from the reduction of overall travel time for users by the development of a project. Because demand for the planned infrastructure is predicted on the basis of VOT, the estimation of VOT is a key element for the determination of the global modal split. This study analyzed the influence of VOT on the profitability of railway projects. Specifically, the components affected by VOT were studied according to their presentation in a railway project appraisal manual. The analysis was performed from two points of view. First, a theoretical formulation was deduced for the relation between VOT, which was included in the cash flow of the different components identified, and the internal rate of return, which represented the profitability of an infrastructure project. Second, a practical case for a suburban railway line was studied to identify the sensitivity of the internal rate of return to VOT, that is, the influence of VOT on the profitability of the project.

Transport project appraisal consists of the quantification of the generalized benefits of a transport project to society, including the gain for transport users, operators, and industry as well as the benefits for nonusers of transport, who gain from the reduction of external effects (noise, air pollution, or climate change, among others), reduction of accident costs, and recovery of productivity lost because of congestion.

The value of time (VOT) is one of the most important elements in project appraisal. VOT is essential for the calculation of the gain derived from time savings for users of new transport infrastructure. This includes the shorter traveling time offered by the improved transport mode and the reduction of time spent in congestion in the alternative modes because of the derived demand from those modes to the improved one. The ex ante prediction of demand for the planned transport infrastructure is also made on the basis of VOT and, consequently, its estimation is a key element for the determination of the global modal split.

VOT can be defined as the subjective valuation that a transport user assigns to the time spent traveling with a certain mode of transport. In neoclassic microeconomics VOT is defined as the willingness to pay for unit travel time savings (*1*).

Most studies dealing with VOT focus on its valuation and estimation. Several factors affecting VOT have been identified. However, less attention has been paid to the influence of its variation on project profitability and, therefore, to the need for a precise estimation of its value.

The present study focused on VOT in the framework of railway infrastructure project appraisals, determining its weight in the overall cost–benefit analysis and, in particular, in the profitability of railway projects. The sensitivity of the internal rate of return (IRR) to VOT is analyzed to determine the necessity of accurate estimation.

This study contributes to the identification and analysis of VOT influence in the profitability of railway infrastructure projects. The analysis is performed from two complementary points of view: a theoretical one, based on the mathematical formulation for cost–benefit analysis calculation, and a practical one, based on a real railway project.

The paper is structured in five sections. The next section reviews the state of the art of VOT estimation in infrastructure projects with a focus on railway projects. VOT importance in project appraisal is emphasized, the different estimation methods are presented, and several values are analyzed. The third section studies the influence of VOT in project appraisal in more detail. The analysis is based on RAILPAG guidelines for railway project appraisal developed by the European Investment Bank as a harmonized tool for project quality analysis in order to compare different railway projects and prioritize investments. The fourth section analyzes the VOT influence in a real case in Barcelona, Spain. The extension of a suburban railway line is studied, and the variation of IRR for the project depending on VOT is analyzed. Finally, the fifth section summarizes the paper's conclusions.

### VALUE OF TIME IN TRANSPORT INFRASTRUCTURE APPRAISAL

VOT was revealed as an important element in transportation appraisals as early as the 1960s. Beesley realized that time savings can represent a high percentage of the generalized benefits to society obtained from the construction of infrastructure projects (*2*). He estimated that time savings accounted for 64% to 78% of the total measured gross benefits (depending on the time unit valuation) in the first operation year of the M1 motorway in the

Center for Innovation in Transport, C. Jordi Girona 29, 2-A, Building Nexus II, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain. Corresponding author: O. Criado, oscar.criado@upc.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 86–94.  
DOI: 10.3141/2261-10

United Kingdom and as much as 80% for the Victoria Line in the London Underground.

These early results emphasized the great weight of time valuation in the calculation of transport benefits to society. In fact, time is a main criterion for the development and planning of transport infrastructures, as these are planned to facilitate travel at the lowest cost in time and money. Furthermore, VOT is essential in the ex ante prediction of the future demand for a transport infrastructure project because it is a key factor in the generalized cost valuation of time spent in each mode of transport. It is a significant parameter to travel behavior analysis and traffic assignment and is, therefore, a key element for the determination of modal split.

Because theoretical studies dealing with VOT have concentrated more on its concept than on its variation mechanisms (1), the factors affecting VOT are well known, but knowledge of how they affect the variation of the value is limited.

VOT theory has been developed in a two-stage process (1). Until the 1960s the utility function for users only included goods consumption. In the first stage of the development of the theory, different authors added the value of nonworking time, work time, and travel time to goods consumption. In 1965 Becker first included nonwork time in the function, which became the basis of the wage-rate approach (3). One year later, Johnson added work time with the consideration that it can be pleasant or unpleasant (4). Some years later, Oort included travel time, considering the benefits obtained through the reduction of travel time (spending that time at work or reducing travel discomfort) because of the construction of a new infrastructure (5).

On the basis of these studies, De Serpa produced a more complete VOT theory, classifying VOT in three types: as a resource (VTR; Equation 1), as a commodity (VTC; Equation 2), and as savings (VTS; Equation 3) (6):

$$VTR = \frac{MU(\mu)}{MU(\lambda)} \quad (1)$$

$$VTC = \frac{MU\left(\frac{\partial u}{\partial t}\right)}{MU(\lambda)} \quad (2)$$

$$VTS = \frac{MU(k)}{MU(\lambda)} \quad (3)$$

where

MU = marginal utility,

$\mu$  = available time,

$\partial u / \partial t$  = time used for an activity,

$\lambda$  = income, and

$k$  = factor related to the limitation of the consumed time.

The second stage in the development of VOT theory entailed a further step in the definition of VOT within the utility function. It consists basically of the analysis of the influence of the use of travel time savings in other activities. Vickrey discussed the impact of arriving at the destination at the desired time on traffic congestion (7). Later, Small introduced a framework for reliability evaluation and deducted that VTR is dependent on the work schedule (8). However, it was Evans who reflected the relationship of time and goods in the income constraint by introducing direct financial cost per hour of time spent in activities (9).

The combination of the models of De Serpa (6) and Evans (9) by Jara-Díaz (10) let Jiang and Morikawa (1) conclude that

- As a necessary input element for activities, time has to be included in the utility function;
- As a limited and uncontrolled resource, the time restriction is imposed on every activity; and
- As a dependent consumption, the time consumption is combined with and is restricted by the goods consumption.

More recently, research in VOT theory has determined that users of transportation systems assign separate values to travel time and reliability of service; the latter is defined as a measure of travel time variation. Small et al. analyzed different studies on VOT estimation for congested and uncongested situations, deducing that congested travel time is valued more than uncongested travel time (11). Lam and Small showed evidence of this variation based on California SR-91 data, determining the relation between the average wage rate and VOT (72%) and the value of reliability for men (48%) and women (101%) (12). Brownstone and Small noted that reliability is valued highly in California, and although a substantial heterogeneity in values across the population was detected, it was difficult to isolate its exact origins (13).

## Estimation Methods for Value of Time

Time is a nonmarket resource, that is, its economic value—how much people would be willing to pay for it—is not revealed in market prices and, therefore, it is not traded in markets. Other nonmarket resources are related to environmental effects, such as air pollution or noise reduction. The monetary values assigned to nonmarket resources may be implicitly undervalued. There are two groups of methodologies for VOT estimation:

- Stated preference and
- Revealed preference.

The stated preference method is an economic technique for the valuation of nonmarket resources based on survey data. It is often used for the valuation of environmental impacts as well as for determining VOT according to different users. Stated preference methods identify the amount of money that people would be willing to pay or be willing to accept for a reduction of time by one unit.

De Borger and Fosgerau classify the commonly used valuation measures in two general groups (14):

- Choices that may be used to reveal compensating variations. The choice is made by comparing the time that it takes to make a journey with the reference mode and the alternative mode, differentiating between the willingness-to-pay measure, when the alternative is faster but more expensive than the reference, and the willingness-to-accept measure, when the alternative is slower but less expensive than the reference.
- Choices that may be used to reveal equivalent variations. The choice is made by comparing the price for making a journey with the reference mode versus the alternative mode. The choice differentiates between the equivalent gain measure, when there is a cheaper alternative that has the same travel time as the reference and another alternative that is faster but with the same cost as the reference, and the equivalent loss measure, when there is a more

expensive alternative that has the same travel time as the reference and another alternative that is slower but with the same cost as the reference.

The revealed preference method is a technique for the valuation of nonmarket resources based on consumer behavior. Consumers' preferences are determined on the basis of their purchase habits. This technique takes into consideration the diminishing rate of substitution, which considers that consumers decide on consumption that maximizes utility.

Various authors have outlined the advantages of adopting stated preference methods to collect survey data over revealed preference techniques. Kroes and Sheldon identified some problems in revealed preference methods, such as the difficulty of obtaining sufficient variation to examine all variables of interest, strong correlations between explanatory variables of interest that make it difficult to estimate model parameters, and the inability to evaluate demand under conditions that do not yet exist (15).

### Variability of Value of Time

VOT estimation is a significant issue for the valuation of the profitability of transport infrastructure projects. VOT estimation methods include different factors in its valuation, and a review of different estimated VOTs in the literature reveals heterogeneity not only for different locations and situations, but even in the same location.

Table 1 shows the great variability in VOT estimation. As the table suggests, the factors influencing VOT estimation include (among others) location (by country and even by region or city) and focus (e.g., by vehicle or travel type). The values also reflect a relationship between VOT and the purchasing power parity per capita of the countries studied: the lower the gross domestic product per capita, the lower the estimated VOT.

These results fit with the conclusions of several authors, who state that VOT cannot be constant, as it depends on a variety of issues, such as travel time, travel fees, and other exogenous factors (e.g., wage rate or work time) (1, 23).

### PROBLEM APPROACH

Within the context described above it is interesting to analyze the VOT effect in project appraisals, identifying the main mechanisms of its influence on determining the profitability of a railway project. The importance of accurately estimating VOT in order to obtain a reliable evaluation of the investment is an interesting component of the analysis as well. A double approach is taken in this study: (a) a theoretical development of the relationship between IRR as a measure of project profitability and VOT (discussed in the present section) and (b) a case study of a real rail line (presented in the following section).

Transportation project appraisal methodologies have been developed in many ways for each transport mode and even in each country. Table 2 reflects the diversity of manuals available in Europe and the United States. One of these, RAILPAG, has recently been developed by the European Investment Bank to provide the European railway sector with a tool for project quality analysis. RAILPAG presents some advantages in comparison with other appraisal manuals (24):

- Alternative modes of transport for the railway project appraisal are considered, in contrast to most rail appraisal manuals, which consider only rail alternatives;
- The socioeconomic analysis is made from the point of view of society as a whole;
- Environmental aspects are considered; and
- A financial quality analysis can easily be made: results are presented in a matrix form (Figure 1) composed of the stakeholders involved and the effects derived in such a way that the socioeconomic result can be studied for each stakeholder–effect binomial relation.

Of the effects derived from a railway project as presented in the RAILPAG matrix, VOT basically affects the users and, in particular, three elements: the value of the time used for traveling, the reliability of service, and the consumer surplus. VOT influence on the profitability of a railway project is studied in the following sections relative to

TABLE 1 Values of Time for Scenarios According to Different Authors

Author	Location	Focus	VOT (€/h) (adjusted to December 2009)	Country GDP per Capita <sup>a</sup> (\$)
Jaro-Arias and Hunt (1996) (16)	Barcelona, Spain	Car	35.1–130.5	33,700
González-Savignat (1999) (17)	Barcelona	Average values	10.7	
Asensio and Matas (1999) (18)	Barcelona	Average values	15.4	
Barrios and Martínez (1999) (19)	Cadiz, Spain	Travels to work	1.5	
Roth and Villoria (2001) (20)	Philippines	Car	0.94	3,300
		Utility vehicle	0.38	
		Jeepney	0.13	
		Bus	0.14	
		Truck	0.80	
		Motorcycle	0.38	
Smalkoski and Levinson (2005) (21)	San Diego, Calif. California	Automobile	49.51	46,400
		Automobile	34.28	
Nash (2010) (22)	United Kingdom	Leisure	8.07	35,200
		Commuting	9.11	
		Business	72.27	

NOTE: GDP = gross domestic product.

<sup>a</sup>GDP value from *The World Factbook*, U.S. Central Intelligence Agency, 2009.

TABLE 2 Review of Manuals for Transportation Project Appraisal in the United States and Europe

Country	Author	Year	Manual
United States	Office of Management and Budget	1992	<i>Guidelines and Discount Rates for Cost–Benefit Analysis of Federal Programs</i>
United States	FAA	1999	<i>FAA Airport Cost–Benefit</i>
Europe	Eurocontrol	2000	<i>Guidelines for the Economic Appraisal of EATMP Projects</i>
Germany	Federal Ministry of Transport, Building, and Housing	2003	<i>Macroeconomic Evaluation Methodology</i>
France	Ministère de l'Équipement, des Transports, du Logement, du Tourisme et de la Mer	2004	<i>Instruction Cadre Relative aux Méthodes d'Évaluation Économique des Projets de Transport</i>
European Union	European Commission. Directorate General for Energy and Transport	2006	<i>HEATCO: Developing Harmonised European Approaches for Transport Costing and Project Assessment</i>
United Kingdom	Department for Transport	2006	<i>Economic Assessment of Road Schemes: The COBA Manual</i>
Europe	European Investment Bank	2007	<i>RAILPAG: Railway Project Appraisal Guidelines</i>

NOTE: EATMP = European Air Traffic Management Program.

			STAKEHOLDERS									
			Users		Operators		Insurance Companies	Infrastructure Administrators		Providers and Contractors	Non-users	Public Administrations (local, regional, national, supranational)
			Rail	Alternative Modes	Rail	Alternative Modes		Rail	Alternative Modes			
EFFECTS	On the users	Tariff										
		Time used										
		Reliability										
		Comfort										
		Security										
		Consumer surplus										
	On the operation	Direct (operation – vehicles, personnel, installations)										
		Indirect (central installation, subsidies, energy taxes, taxes)										
	On the assets	Investment (infrastructure, superstructure, rolling stock ...)										
		Maintenance (infrastructure, rolling stock)										
	External	Air pollution										
		Noise										
		Climate change										
		Use of space										

FIGURE 1 RAILPAG matrix of stakeholders and effects in a railway project.

these three elements. Although differences in the valuation of travel time and reliability of service have been identified in several works, in the present paper VOT is considered uniform for both elements as the relationship between their values is not yet well defined.

The theoretical approach is based on RAILPAG guidelines and the IRR function. IRR is the ratio of money gained or lost on an investment relative to the amount of investment money used. It is commonly used for measuring and comparing the profitability of investments. For transportation project appraisals, the cost–benefit analysis includes user costs and benefits, and therefore it is a proper basis for social profitability calculation as it includes both economic and noneconomic aspects.

The formula for the IRR calculation, based on net present value, is expressed as

$$\sum_{i=1}^n \frac{V_i}{(1 + \text{IRR})^i} - I_0 = 0 \quad (4)$$

where

$V_i$  = cash flow in period  $i$ ,  
 $I_0$  = initial investment, and  
 $n$  = number of years.

For the theoretical approach, a formula is deduced on the basis of Equation 4 that relates profitability and VOT. The components of the effects on users that are identified as VOT-dependent (VOT used for travel, reliability of service, and consumer surplus) are explicitly defined. The socioeconomic analysis of a transportation project proposed by RAILPAG guidelines is made from the point of view of society as a whole, and therefore considers the effects on users, on operation, and on assets as well as external effects, including both economic and noneconomic aspects, in order to make them more understandable and to allow a comparison of their impacts.

From Equation 4, two developments ensue:

- Deduction of a VOT-dependent function for the cash flow calculation and
- Rewriting of a clearer expression of the relationship between IRR and cash flow, and consequently between IRR and VOT.

For the first development, total cash flow in period  $i$  ( $V_i$ ) is the sum of the cash flow of each effect derived from the transport infrastructure project. Its calculation on the basis of RAILPAG guidelines can be written as

$$V_i = \text{CB}_{\text{us}_i} + \text{CB}_{\text{op}_i} + \text{CB}_{\text{as}_i} + \text{CB}_{\text{ee}_i} \quad (5)$$

where

$\text{CB}_{\text{us}_i}$  = cost–benefit of the effects on users,  
 $\text{CB}_{\text{op}_i}$  = cost–benefit of the effects on operation,  
 $\text{CB}_{\text{as}_i}$  = cost–benefit of the effects on assets, and  
 $\text{CB}_{\text{ee}_i}$  = cost–benefit of external effects.

Taking into account the influence of VOT described above, the last three components of Equation 5 are not VOT-dependent and can be considered constant for the analysis. The formula can be transformed as

$$V_i = \text{CB}_{\text{us}_i} + \delta_i \quad (6)$$

with

$$\delta_i = \text{CB}_{\text{op}_i} + \text{CB}_{\text{as}_i} + \text{CB}_{\text{ee}_i}$$

where  $\delta_i \neq f(\text{VOT})$ .

Consequently, the analysis of VOT influence on the profitability of a railway project focuses on the cost–benefit of the effects on users. According to RAILPAG guidelines, this component includes VOT used for traveling, reliability of service, consumer surplus, and comfort and security. Comfort and safety are not valued based on VOT and, therefore, can be considered constant in the analysis. VOT used for traveling is defined as the value of the total amount of time used by all users, and reliability of service is defined as the value of the total amount of time spent in transport delays. The further expression for the cost–benefit calculation of the effects on users can then be written as

$$\text{CB}_{\text{us}_i} = \text{VTS}_i + \text{RS}_i + \text{CS}_i + \gamma_i + \beta_i \quad (7)$$

where

$\text{VTS}_i$  = value of the time used for traveling,  
 $\text{RS}_i$  = reliability of service,  
 $\text{CS}_i$  = consumer surplus,  
 $\gamma_i = \text{com}_i \neq f(\text{VOT})$  for comfort, and  
 $\beta_i = \text{saf}_i \neq f(\text{VOT})$  for safety.

With Equation 7 introduced into Equation 6, the cash flow for the period  $i$  ( $V_i$ ) is calculated as

$$V_i = \text{VTS}_i + \text{RS}_i + \text{CS}_i + \alpha_i \quad (8)$$

where  $\alpha_i$  is equal to  $\beta_i + \gamma_i + \delta_i$ .

This calculation is developed for each competitor mode considered. A reference transport mode is defined as a basis for a comparative evaluation of costs and benefits of the rest of the competitor modes. In a railway project, the reference mode is the train mode, and the alternative modes include other railway services, urban and interurban bus services, truck services, and automobiles.

Equation 8 can be adapted for each transport mode in competition. The calculation of VTS and RS for transport mode  $j$  (automobile, bus, or train) in time period  $i$  can be expressed as

$$\text{VTS}_{ij} = \text{tt}_{ij} \times v_{ij} \times u_{ij} \times \text{VOT}_{ij} \quad (9)$$

$$\text{RS}_{ij} = \text{ad}_{ij} \times \text{vwd}_{ij} \times v_{ij} \times u_{ij} \times \text{VOT}_{ij} \quad (10)$$

where

$\text{tt}_{ij}$  = travel time,  
 $v_{ij}$  = number of vehicles per time unit traveling,  
 $u_{ij}$  = number of users per vehicle,  
 $\text{ad}_{ij}$  = average delay of vehicles arriving late, and  
 $\text{vwd}_{ij}$  = percentage of vehicles with delay.

On the basis of these definitions, the expression in period  $i$  for transport mode  $j$  for the cash flow calculation is rewritten as

$$V_{ij} = (\text{tt}_{ij} + \text{ad}_{ij} \times \text{vwd}_{ij}) \times v_{ij} \times u_{ij} \times \text{VOT}_{ij} + \text{CS}_{ij} + \alpha_{ij} \quad (11)$$

where  $V_{ij}$  is cash flow and  $\alpha_{ij}$  is a VOT-nondependent term.

The consumer surplus in period  $i$  for transport mode  $j$  ( $CS_{ij}$ ) is defined as the utility; after conversion to a monetary value that a person receives in the choice situation (25),  $CS_{ij}$  is calculated with

$$CS_{ij} = (GC_{ij} - GC_i^{\text{ref}}) \times v_{ij} \times u_{ij} \quad (12)$$

where  $GC_{ij}$  is the generalized cost, and  $GC_i^{\text{ref}}$  is the generalized cost for the reference mode.

$GC_{ij}$  is calculated as the sum of the average tariff ( $AT_{ij}$ ) and the VOT spent for traveling ( $VOT_{ij}$ ), including travel time ( $tt_{ij}$ ), access time ( $at_{ij}$ ), and waiting time ( $wt_{ij}$ ):

$$GC_{ij} = AT_{ij} + VOT_{ij} \times (tt_{ij} + 2at_{ij} + wt_{ij}) \quad (13)$$

As a result, the following expression is obtained for the cash flow calculation for transport mode  $j$  in period  $i$  ( $V_{ij}$ ):

$$V_{ij} = v_{ij} \times u_{ij} \times \left( AT_{ij} - GC_i^{\text{ref}} + \left( 2tt_{ij} + 2at_{ij} + wt_{ij} \right) \times VOT_{ij} \right) + \alpha_{ij} \quad (14)$$

Including this result in Equation 4, the next equation can be deduced for  $n$  periods considering the Pascal triangle and the Newton binomium:

$$I_0 \times \sum_{k=0}^n \left[ \frac{n!}{k!(n-k)!} \times IRR^{(n-k)} \right] - \sum_{i=1}^n \left[ V_i \times \sum_{k=0}^{n-1} \left[ \frac{(n-i)!}{k!(n-i-k)!} \times IRR^{(n-i-k)} \right] \right] = 0 \quad (15)$$

where  $n$  is the number of years of the study period, and  $V_i$  is the cash flow in period  $i$  ( $i = 1, 2, \dots, n$ ).

The solution of Equation 15 is not trivial and it is not the aim of this paper to solve it. Further research is under way by the authors to discern if there is a real solution to the equation. However, taking into account Equation 14 for the cash flow calculation of each transport mode, the following conclusions can be made:

- The existence of a relationship between the profitability of a railway project measured with IRR and VOT is confirmed and
- Although VOT influences the annual total cash flow ( $V_i$ ) linearly, the IRR variation with VOT seems to be more complex, as it is the solution of an  $n$ -dimensional equation.

## CASE STUDY: EXTENSION OF SUBURBAN RAILWAY LINE IN BARCELONA

This case study analysis is based on RAILPAG guidelines. A socioeconomic appraisal for a real project was performed, and the sensitivity of IRR to VOT was analyzed.

The case study consists of the extension of a suburban railway line affecting two mid-sized cities (Cities A and B) of 200,000 inhabitants in the metropolitan area of Barcelona. The line is composed of a trunk section and two branches. Each city has a branch con-

necting to the trunk line, which in turn connects to the larger city of Barcelona. The extension project in question consists of the development of a double-track tunnel infrastructure within the urban area of each city: a 4.5-km length and three new stations in the case of City A, and a 5.3-km length and four new stations in the case of City B (Figure 2a).

Competitor modes are defined separately for intracity journeys inside Cities A and B and for the intercity journeys between the two cities and the center of Barcelona. The train service on the extended section is identified as the reference mode for both journey types. The alternative modes are identified separately: for intracity journeys, automobile and an urban bus service, and for intercity journeys, automobile and an existing alternative railway mode. Services of the public transport modes are detailed in Figure 2b, and the data input used in the case study for the sensitivity analysis is shown in Figure 2c.

The results of the test of IRR sensitivity to the variation of VOT are represented in Figure 3, in which the variation of a unitary IRR value is related to a unitary VOT. The unitary VOT is defined as the relationship between VOT and a VOT of reference ( $VOT/VOT_{\text{ref}}$ ).  $VOT_{\text{ref}}$  was a conservative estimate based on published values for the study scenario (€6.8 for public transport modes and €8.7 for automobile; €1 = \$1.22; all euro values are for 2004). The unitary variation of IRR is defined as the relationship between the variation of IRR and an IRR reference value ( $\Delta IRR/IRR_{\text{ref}}$ );  $IRR_{\text{ref}}$  was calculated on the basis of  $VOT_{\text{ref}}$ .

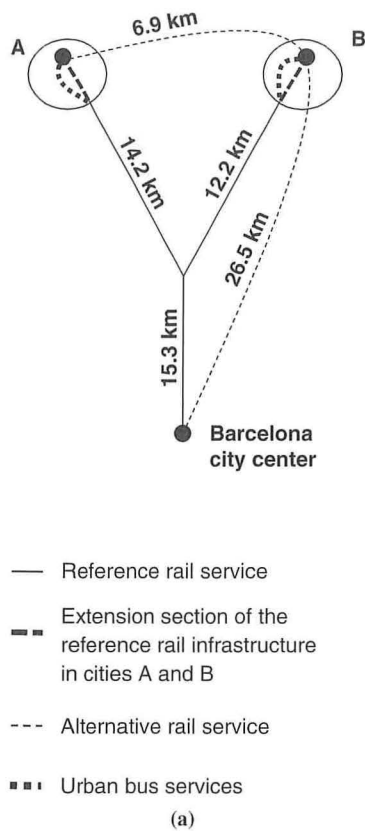
As Figure 3 shows, the sensitivity of IRR to VOT depends on the mode of transport. In fact, the results differ for the reference mode and the alternative modes.

Concerning the reference mode, the global scenario reflects the following:

- VOT has a great influence on project profitability. This finding is explained by the influence of this value on the cash flow calculation. VOT is taken into account in the calculation of the three components influenced by VOT (value of travel time, reliability of service, and consumer surplus) of the reference mode itself and in the consumer surplus calculation of the alternative modes.
- IRR variation is inversely proportional to the reference mode. This relationship is explained by the negative influence of its value on the consumer surplus of the alternative modes. As an example, the increase of the reference mode VOT implies that the generalized cost of this mode increases. Therefore, the consumer surplus of the alternative modes calculated on the basis of the difference between the generalized cost of the alternative mode and the generalized cost of the reference mode (Equation 12) decreases. This is translated into a lower gain for the users of the alternative mode, decreasing their motivation to change mode. The fewer the users changing from the alternative modes to the reference one, the lower the project profitability.

The influence on project profitability of the respective VOTs of the alternative modes is lower than that of the reference mode. Nevertheless, this influence is not negligible. From the results of Figure 2, the following can be pointed out:

- A 2% VOT increase of those modes supposes an IRR increase of 2.6% in the urban bus case, 2.4% in the automobile case, and 1.3% in the alternative train case.



MODE OF TRANSPORT	OPERATIONAL ISSUES FOR PUBLIC TRANSPORT	
Reference rail mode	Current operation	12-minute frequency service between the two cities and Barcelona city center
	Planned operation	6-minute frequency service inside the two cities
Alternative rail service	Both cities are served with the same rail service	
	Current and future Operation	14-minute frequency service
Urban bus	City A	5 lines serving end points of the extension
	City B	11 lines serving end points of the extension

(b)

COMPONENT		JOURNEY	VALUE IN THE FIRST YEAR				ANNUAL GROWTH RATE (subsequent 29 years)
			Reference rail mode	Urban bus	Alternative rail mode	Car	
Average tariff	€/user	AT <sub>ij</sub>	Intra-urban	0.58	-	2.48 <sup>(1)</sup>	+3%
			Inter-urban	-	1.02		
Travel time	h	tt <sub>ij</sub>	Intra-urban	0.503	-	0.20	Constant
			Inter-urban	-	0.519	0.55	
Access time	h	a <sub>ij</sub>	Intra-urban	0.05	-	0.03	Constant
			Inter-urban	-	0.12	0.05	
Waiting time	h	wt <sub>ij</sub>	Intra-urban	0.12	0.60	-	Constant
			Inter-urban	-	0.12	0	
Average delay	h	ad <sub>ij</sub>	Intra-urban	0.10	0 <sup>(2)</sup>	-	Constant
			Inter-urban	-	0.17	0 <sup>(2)</sup>	
Vehicles with delay	%	vwd <sub>ij</sub>	Intra-urban	0.25	0 <sup>(2)</sup>	-	Constant
			Inter-urban	-	3.17	0 <sup>(2)</sup>	
Vehicles	Veh./day	v <sub>ij</sub>	Intra-urban	155	915	-	0% <sup>(3)</sup> for public transport
			Inter-urban	-	-	154	+1.45% for car
Users	User/veh.	u <sub>ij</sub>	Intra-urban	151	40	-	+2% for public transport
			Inter-urban	-	-	128	0% for car

Notes:

<sup>(1)</sup> Average tariff for the car is calculated on the basis of the average fuel consumption, value of time spent looking for parking, and parking cost

<sup>(2)</sup> Conservative estimate because of lack of data

<sup>(3)</sup> No increases in frequency for public transport

(c)

FIGURE 2 Characteristics of the study case: (a) extension of suburban rail line in Barcelona, (b) operational issues for public transport modes, and (c) data input for case study.

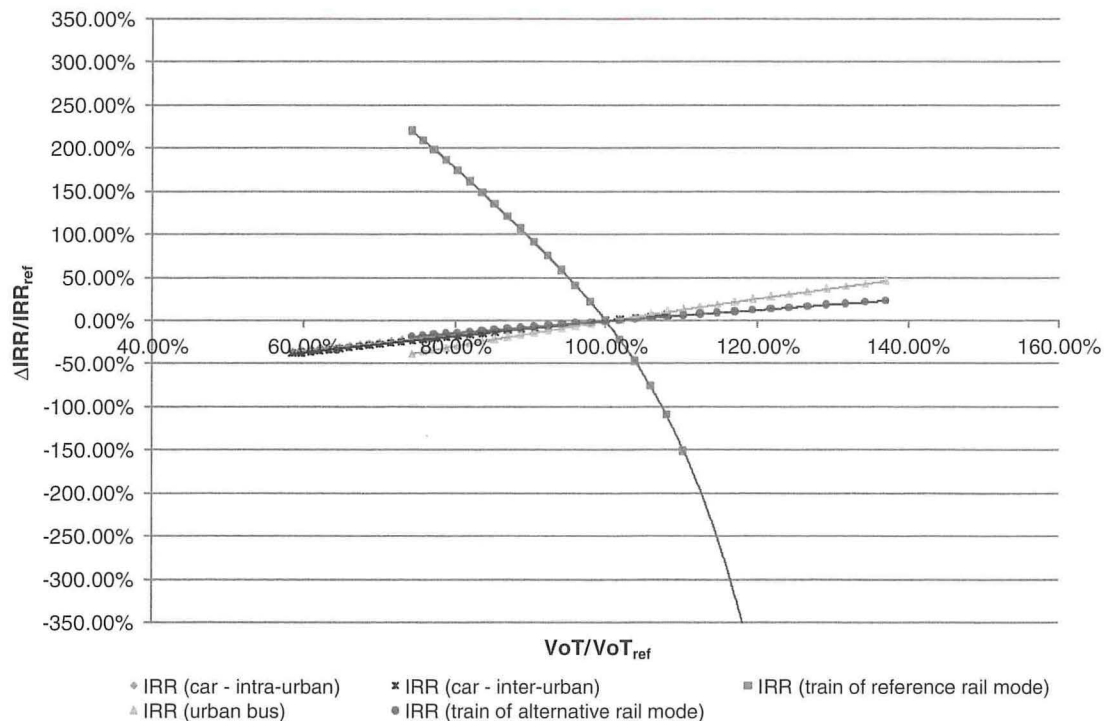


FIGURE 3 Variation of IRR with VOT for global scenario (including extension of line in both medium cities).

**TABLE 3** Difference in Generalized Cost Between Reference Train Mode and Alternative Modes

Journey Type	Reference Train Versus Urban Bus (€)	Reference Train Versus Automobile (€)	Reference Train Versus Alternative Train (€)
Intraurban	2.77	0.39	—
Interurban	—	2.63	2.20

NOTE: — = not applicable.

• The variation of VOT of the alternative modes only affects the respective cash flow ( $V_{ij}$ ) element and, therefore, it is consistent that the influence of these values in the project IRR is lower than that of the reference mode. In fact, the competition level of the alternative mode in relation to the reference mode determines the influence of its VOT in project profitability. This fact is reflected by the differences in the generalized costs shown in Table 3. The bigger the difference between the generalized costs of two modes, the more competition there is. Consequently, in intracity journeys the reference mode is more competitive than the alternative public transport mode, whereas in interurban journeys it is more competitive than the automobile. Figure 2 confirms that the alternative mode whose VOT has more influence in the IRR of the project is the urban bus for intracity journeys and the automobile for intercity ones.

Although VOT influence on project profitability depends on the mode of transport, the sensitivity of the IRR of the project can also vary for the same mode of transport. This can be observed when comparing the alternative public transport offered in both cities with the sensitivity of the project profitability (Figure 4).

The results show that the availability of a better public transport alternative means a reduced influence of VOT on project profitability. Indeed, the alternative public transport is notably better in City B than in City A, both for the urban bus in the intracity journeys (11 lines in City B compared with five lines in City A) and for the alternative train in the intercity journeys (City B is closer to the Barcelona city center, which means a shorter trip).

## CONCLUSIONS

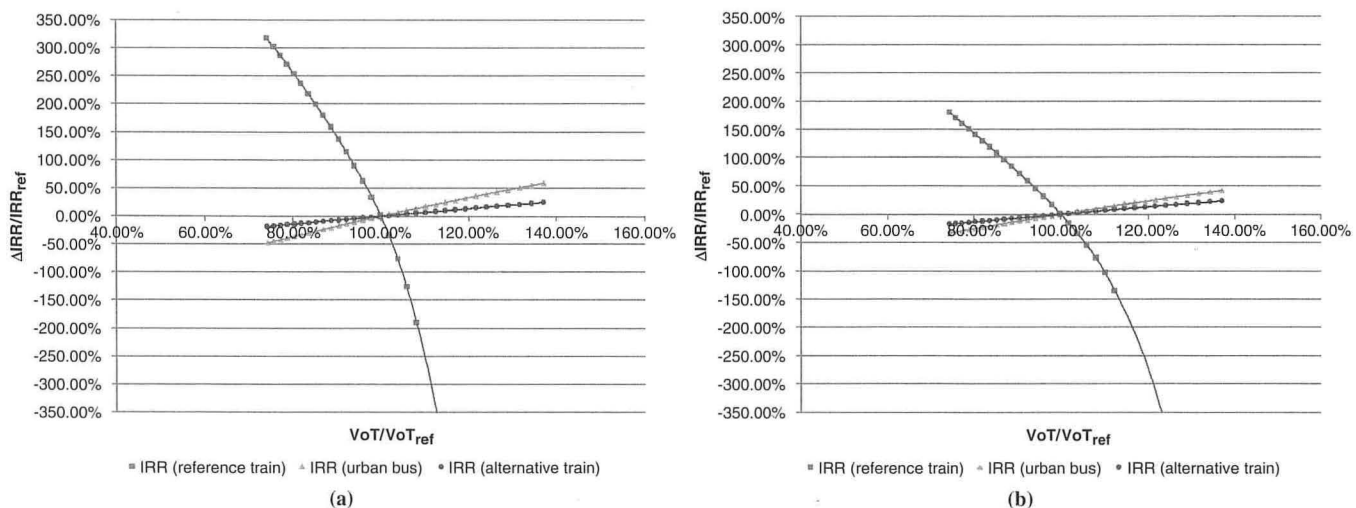
This study considered the influence of VOT on the profitability of railway projects in two ways. A theoretical study analyzed project appraisal for railway infrastructures, particularly those components affected by VOT, and a practical case of the extension of a railway line in two medium-sized cities was studied to identify how VOT variation affects project profitability.

In the theoretical study, a formulation was deduced relating IRR, the parameter representing project profitability, and VOT, deduced in the cash flow of the three components considered by RAILPAG guidelines in a railway project appraisal: valuation of travel time, reliability of service, and consumer surplus. It can be concluded that a relation exists between project profitability and VOT. Although VOT linearly influences annual total cash flow ( $V_i$ ), IRR variation with VOT seems to be more complex, as it is the solution of an  $n$ -dimensional equation. Therefore, further research is being carried out by the authors to determine if there is a real solution possibility to the equation deduced.

The practical study considered the influence of VOT variation on IRR for the case of an extension of a suburban railway line in two medium-sized cities in the metropolitan area of Barcelona. The results reflect the following:

- VOT has a great influence on profitability in the case study analyzed. This influence is more important for the reference mode.
- IRR sensitivity to VOT is greater when the competition of the alternative modes is greater. For the case study, the greatest sensitivity is to the urban transport VOT in the intracity journeys and to the automobile VOT in the intercity journeys.
- The better the alternative transport offer, the lower the profitability sensitivity of the project to VOT. That is to say, the better the improvement of the reference mode regarding the alternative modes, the higher the sensitivity of IRR to VOT.

From the practical study it can be concluded that it is necessary to estimate VOT precisely in order to achieve a reliable economic evaluation of a railway project. If this estimation cannot be carried out, it is recommended that a conservative estimation of VOT be used in order to avoid overestimating project profitability.

**FIGURE 4** Variation of IRR with VOT for case study for extension project in (a) City A and (b) City B.

## REFERENCES

1. Jiang, M., and T. Morikawa. Theoretical Analysis on the Variation of Value of Travel Time Savings. *Transportation Research Part A*, Vol. 38, 2004, pp. 551–571.
2. Beesley, M. E. The Value of Time Spent in Travelling: Some New Evidence. *Economica*, New Series, Vol. 32, No. 126, 1965, pp. 174–185.
3. Becker, G. S. A Theory of the Allocation of Time. *Economic Journal*, Vol. 75, No. 299, 1965, pp. 493–517.
4. Johnson, M. Travel Time and the Price of Leisure. *Western Economic Journal*, Vol. 4, 1966, pp. 135–145.
5. Oort, O. The Evaluation of Traveling Time. *Journal of Transport Economics and Policy*, No. 3, 1969, pp. 219–286.
6. De Serpa, A. C. A Theory of the Economics of Time. *Scottish Journal of Political Economy*, Vol. 19, 1972, pp. 828–846.
7. Vickrey, W. S. Congestion Theory and Transport Investment. *American Economic Review*, Vol. 59, No. 2, 1969, pp. 251–260.
8. Small, K. A. Scheduling of Consumer Activities: Work Trips. *American Economic Review*, Vol. 72, 1982, pp. 467–479.
9. Evans, S. On the Theory of the Valuation and Allocation of Time. *Scottish Journal of Political Economy*, Vol. 19, 1972, pp. 1–17.
10. Jara-Díaz, S. R. On the Goods–Activities Technical Relations in the Time Allocation Theory. *Transportation*, Vol. 30, 2003, pp. 245–260.
11. Small, K. A., R. Noland, X. Chu, and D. Lewis. *NCHRP Report 431: Valuation of Travel-Time Savings and Predictability in Congested Conditions for Highway User-Cost Estimation*. TRB, National Research Council, Washington, D.C., 1999.
12. Lam, T. C., and K. A. Small. The Value of Time and Reliability: Measurement from a Value Pricing Experiment. *Transportation Research Part E*, Vol. 37, 2001, pp. 231–251.
13. Brownstone, D., and K. A. Small. Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations. *Transportation Research Part A*, Vol. 39, 2005, pp. 279–293.
14. De Borger, B. D., and M. Fosgerau. Discrete Choices and the Trade-Off Between Money and Time: A Test of the Theory of Reference-Dependent Preference. *MPRA Paper*, No. 12565, 2007.
15. Kroes, E. P., and R. J. Sheldon. Stated Preference Methods: An Introduction. *Journal of Transport Economics and Policy*, Vol. 22, No. 1, 1988, pp. 1–25.
16. Jaro-Arias, L., and D. Hunt. The Value of Travel Time by Private Vehicle in Barcelona Area. *Review of Transport and Communication Studies*, Vol. 65, 1994.
17. González-Savignat, M. The Value of Time. *Papers of Spanish Economy*, No. 82, 1999, pp. 262–275.
18. Asensio, J., and A. Matas. The Public Transport Role in the Cities. *Papers of Spanish Economy*, Vol. 82, 1999, pp. 178–193.
19. Barrios, M. C., and M. A. Martínez. The Value of Time in Work Travels: An Estimation for Cadiz. *Review of Applied Economy Studies*, Vol. 12, 1999, pp. 5–16.
20. Roth, G., and O. G. Villoria, Jr. Finances of a Commercialized Urban Road Network Subject to Congestion Pricing. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1747, TRB, National Research Council, Washington, D.C., 2001, pp. 29–35.
21. Smalkoski, B., and D. Levinson. Value of Time for Commercial Vehicle Operators in Minnesota. *Journal of the Transportation Research Forum*, Vol. 44, No. 1, 2005, pp. 89–102.
22. Nash, C. Transport Investment Appraisal—European Research and Practice: Rail Project Appraisal with Key Issues and Closure Example. Course on transport appraisal, Technical University of Catalonia, Barcelona, Spain, 2010.
23. Kono, T., H. Morisugi, and A. Kishi. When Is the Concept of Generalized Transport Costs Useless? The Effects of the Change in the Value of Time. *Urban Transport*, Vol. 14, 2008, pp. 629–638.
24. *Research Programme on the Economics Effects of Infrastructure (OEEL)*. CPB Netherlands Bureau for Economic Policy Analysis and Netherlands Economic Institute, Rotterdam, Netherlands, 2002.
25. De Jong, G., A. Daly, M. Pieters, and T. Van der Hoorn. The Long Sum As an Evaluation Measure: Review of the Literature and New Results. *Transportation Research Part A*, Vol. 41, 2007, pp. 874–889.

---

*The Freight Transportation Economics and Regulation Committee peer-reviewed this paper.*

# Expanding Alaska–Canada Rail

## Jointly Visualizing Revenue Freight, Development Cost, Mineral Commodity Value, and Impact of Carbon Dioxide

Colin Brooks, Helen Kourous-Harrigan, Mike Billmire, Paul Metz,  
D. Eric Keefauver, Robert Shuchman, Richard Dobson,  
K. Arthur Endsley, and Mark Taylor

Recent changes in global markets have raised the value of mineral resources in northwestern Canada and Alaska. The development of these resources depends on the economics of rail infrastructure expansion. Transportation decision makers need revenue and cost assessments to plan investment in rail infrastructure. A tool based on a geographic information system was developed for mineral resource evaluation and visualization. The tool incorporated expert-augmented mineral resource data for more than 22,000 occurrences in the region. The tool included the proposed Alaska–Canada Rail Link, which would connect Alaska rail to the lower 48 states. Users selected locations of known mineral occurrences near actual or proposed rail routes and used statistical mineral deposit models to estimate resource sizes and extractable value over time by combining current or user-entered commodity prices with multimodal revenue freight volumes and optimally routed transportation costs. The tool translated the revenue scenario into likely carbon dioxide emissions according to the transport of mineral concentrates to regional and international destinations. Users could select and visualize multimodal transportation networks to understand and minimize mobile-source carbon emissions as part of their scenarios. Statistical estimates of mine capital expenditure and operating costs were also calculated according to type. The tool calculated the gross metal value of a mineral occurrence with statistical deposit models. This index was linked to the positive regional economic impact associated with the developed resource in terms of jobs, taxes and royalties, and resupply. This information helped decision makers close the loop on infrastructure investment assessments.

The mineral occurrence estimation and visualization (MOREV) tool presented in this paper is a geographic information system (GIS)–based tool that provides valuable support for decision makers and stakeholders considering freight rail extensions, mine development, and other related activities in Alaska and Canada.

The development of this tool was motivated by two primary factors. First, changes in global commodity markets have renewed interest

in resource deposits in Alaska and in Yukon and British Columbia, Canada, the development of which depends on investment in new rail links between existing networks. Specifically, although a rail link from Alaska to the rest of the North American rail system has been under consideration since 1916 (1) and the mutually dependent economics of large-scale northern resource and railway development are compelling, the plans for the link have never come to fruition.

From 2005 through 2007, an extensive Phase I feasibility study was performed for the Yukon government and the State of Alaska (1) to determine the extent of a business case for investment in the Alaska–Canada Rail Link (ACRL) as part of an emerging North Pacific Rim trade corridor. In the study, a few proposed mineral resource development or extension projects enabled a revenue freight forecast for iron and coal, mineral concentrates, pipeline construction and resupply, and Asian intermodal container imports for a 50-year project life cycle of \$35 billion and 1,989 million tons (Figure 1). Thus, the proposed rail link is economically promising and potentially economically profitable, depending on assumptions made on the discount rate on investment and the types and numbers of resources that would be developed near the rail. The Phase I report also discusses the likely public investment needed to bring this long rail route to fruition. The new study and the associated GIS tool described in this paper augment and expand on the work done in the Phase I feasibility study by analyzing more mineral commodities than were considered in the Phase I study to provide a more complete and detailed picture of potential mineral revenues and train freight volumes. These developments enable users to perform detailed sensitivity analyses to determine if mineral occurrence development could be part of an economically feasible rail route extension.

The current study, which focuses on a 100-km-wide corridor surrounding the proposed rail link as an area likely to become more accessible if rail links are installed, includes a North Slope spur paralleling the Trans-Alaska Pipeline. However, the developed GIS-based framework is extensible to the entire Alaska and Northern Canada region, or any region for which mineral occurrence data are associated with a mineral deposit model (discussed below).

The second primary motivation for this study and tool is the possibility of national climate change legislation in the near future. In order to incorporate carbon caps or taxes into their business decisions, decision makers in the energy, resource development, and transportation sectors will need tools to estimate the impact of transportation carbon emissions for the life cycle of new projects. The MOREV tool includes a mobile-source carbon emissions calculator that will assist with these needs.

---

C. Brooks, H. Kourous-Harrigan, M. Billmire, D. E. Keefauver, R. Shuchman, R. Dobson, and K. A. Endsley, Michigan Tech Research Institute, Michigan Technological University, 3600 Green Court, Suite 100, Ann Arbor, MI 48105. P. Metz and M. Taylor, Department of Mining and Geological Engineering, University of Alaska, Fairbanks, P.O. Box 755800, Fairbanks, AK 99775-5800. Corresponding author: H. Kourous-Harrigan, helen.harrigan@mtu.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 95–105.  
DOI: 10.3141/2261-11

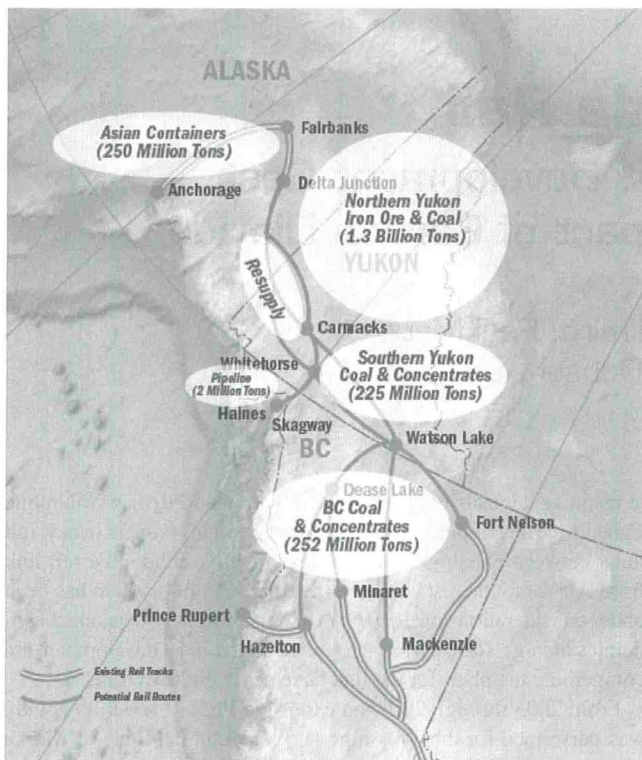


FIGURE 1 Freight forecasts for proposed Alaska-Canada rail link including long-range markets and route options from Phase I feasibility study (BC = British Columbia) (1).

Potential users of the MOREV tool include small- and medium-sized exploration interests in the prefeasibility stages of project planning; transportation and infrastructure planners; state and local government agencies, including those responsible for environmental regulation and permitting; investment community representatives; and researchers. The MOREV tool has wide-ranging application as a decision support tool for these communities and others as it

brings together and jointly visualizes transportation networks, carbon impacts from transportation, commodity development costs and revenues, and revenue freight for rail and intermodal infrastructure investment. The route planning portion could be used in conjunction with and in support of sensitivity analyses generated by capacity evaluation tools such as the Canadian National Railway parametric model (2).

## GENERAL APPROACH

The MOREV tool is a map-based visualization tool for the ACRL region that incorporates potential revenue, traffic volumes, and mobile-source carbon emissions. Central to the tool is a linking between the mineral occurrence database from the Phase I study (1) (developed by P. Metz, professor of geological engineering at the University of Alaska, Fairbanks) and GIS layers of regional mineral occurrences in Alaska, British Columbia, and Yukon (developed by the U.S. Geological Survey and the State of Alaska for the United States and by Canadian territorial and provincial governments for Canada). Linking the mineral occurrence database to these GIS layers allows resource development and transportation decision makers to integrate detailed mineral value data and examine customizable scenarios on the basis of this data.

MOREV is scenario driven; that is, the workflow allows for visualization of the aggregate region, including resources and transport links; downselection to a potential development site; optimization of revenue freight routes, including multimodal shifts and overseas destinations via ocean-going vessels; estimation of resource development and operating costs; and estimation of carbon dioxide (CO<sub>2</sub>) impacts.

The tool allows users to

- Filter occurrences by commodity type, land status, deposit size, or other attribute (Figure 2);
- Calculate potential revenue from occurrences within 100 km of a proposed transport link;
- Visualize proximity to existing infrastructure, historic mines, or nearby deposits;
- Visualize land use patterns, watersheds, and political boundaries;

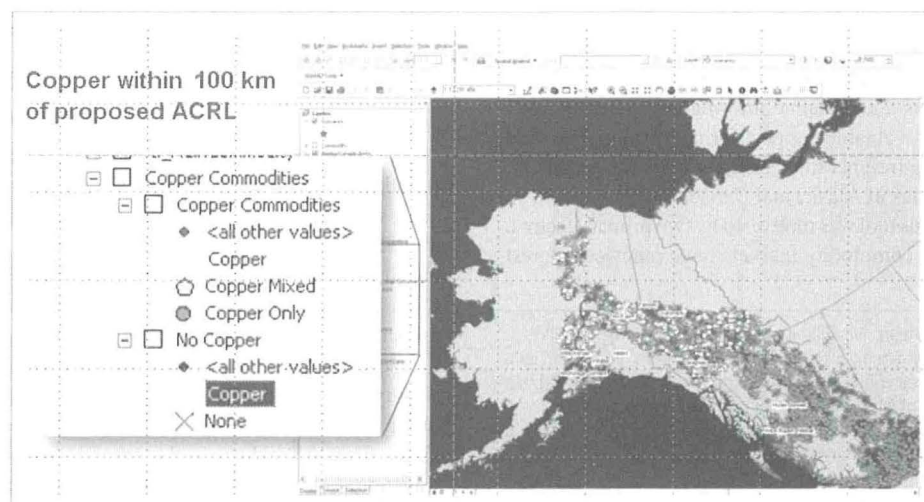


FIGURE 2 MOREV screen shot showing occurrence data filtered by 100-km buffer of proposed rail link and by models containing copper and mixed copper.

- Track CO<sub>2</sub> emissions in the transportation segment for a proposed mine; and
- Calculate and visualize the most efficient multimodal transportation route.

Inputs and assumptions are transparent to and modifiable by the user. Some of the parameters exposed to the user are

- Transport parameters, such as modal shift costs, freight and carbon cost per ton-mile, and port charges (Figure 3);
- Resource parameters, such as deposit tonnage and commodity price; and
- Extraction parameters, such as annual operating cost, mine recovery rate, and extraction rate.

As a result of this capability, various sensitivity analyses can be performed around a given scenario. For example

- Transportation costs can be compared with and without a new rail link;
- Carbon impacts of multimodal routing options can be compared to contrast shifts between heavy haul truck, rail, and ocean-going vessels, as in Milford and Allwood (3);
- Expected revenues over time can be compared through varying mineral commodity prices; and

- Conservative versus generous estimates of metal value can be compared through data for frequency of occurrence from the 10th, 50th, or 90th percentile mineral occurrence tonnage and grade models.

Mineral occurrence data and other parameters can be easily updated as new information is available, thus ensuring the tool's utility for decision support well into the future.

Figures 3 through 5 demonstrate the tool being used to assess a particular mineral occurrence. Figure 3 shows the tool's costing scenario module after selection of a particular occurrence, in this case the Fairplay porphyry copper–molybdenum occurrence north of Tok, Alaska. The potential gross revenue, capital costs to develop the mine, transportation costs, gross metal value (GMV), tonnage, default mine type, and user-selected scenario length features of the MOREV tool can all be seen. Figure 4 shows the ArcGIS interface to the tool's underlying data, including mineral occurrences within 100 km of the proposed rail lines (dark dots), other mineral occurrences (gray dots), the 100-km buffer (yellow line), and the proposed and existing rail lines (dark colored lines). Figure 5 shows the results of a user's selecting the port of Bayuquan, China, as the mineral concentrate destination. The tool's routing functionality automatically selects and displays the shortest shipping, rail, and road routes. That each occurrence tied to a mineral deposit model can be evaluated in this way provides opportunities for users to discover unanticipated insights related to mineral development and rail extension in Alaska.

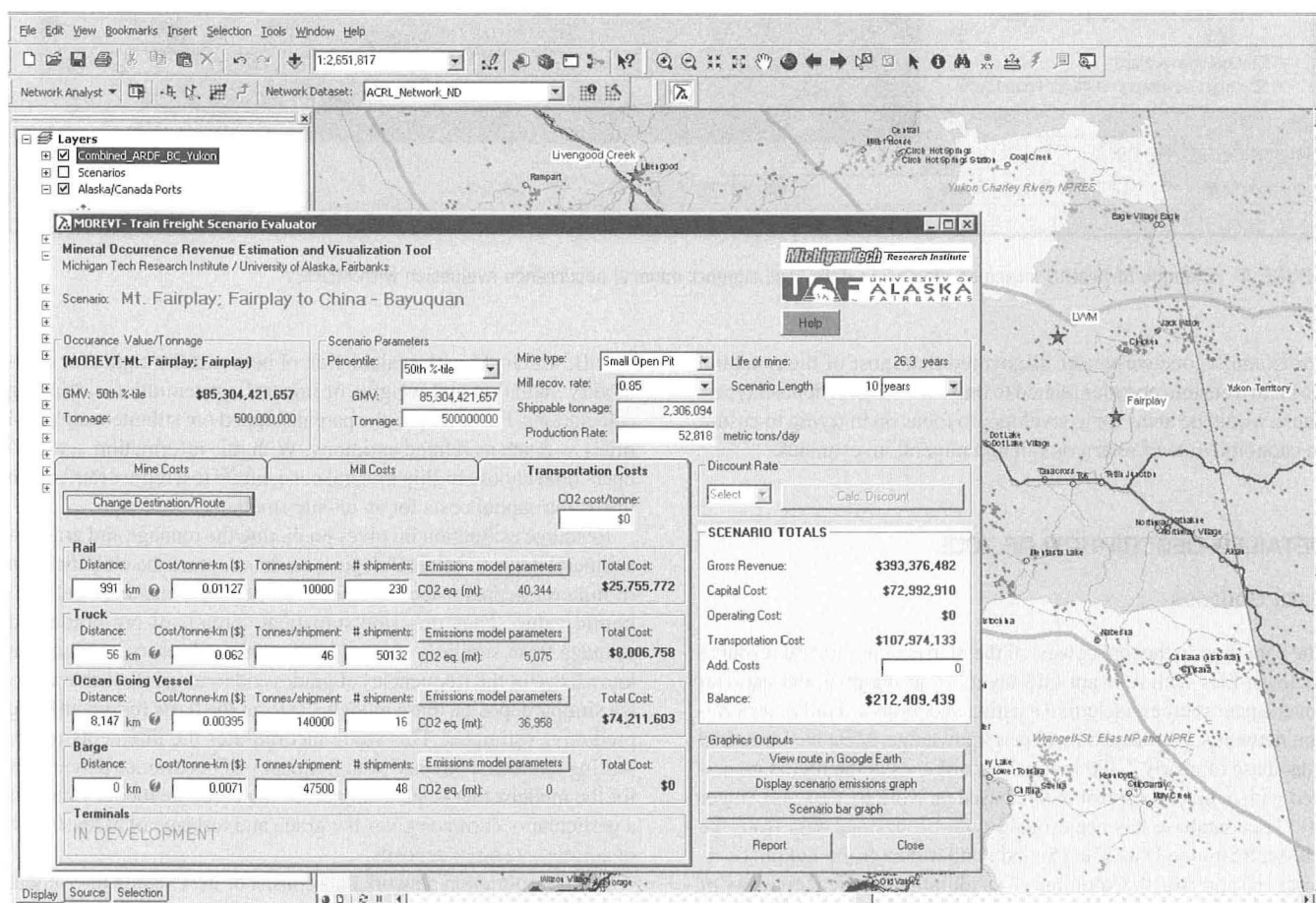


FIGURE 3 Example of MOREV's transportation costing module activated after user selection of Fairplay porphyry copper–molybdenum mineral occurrence for further evaluation.

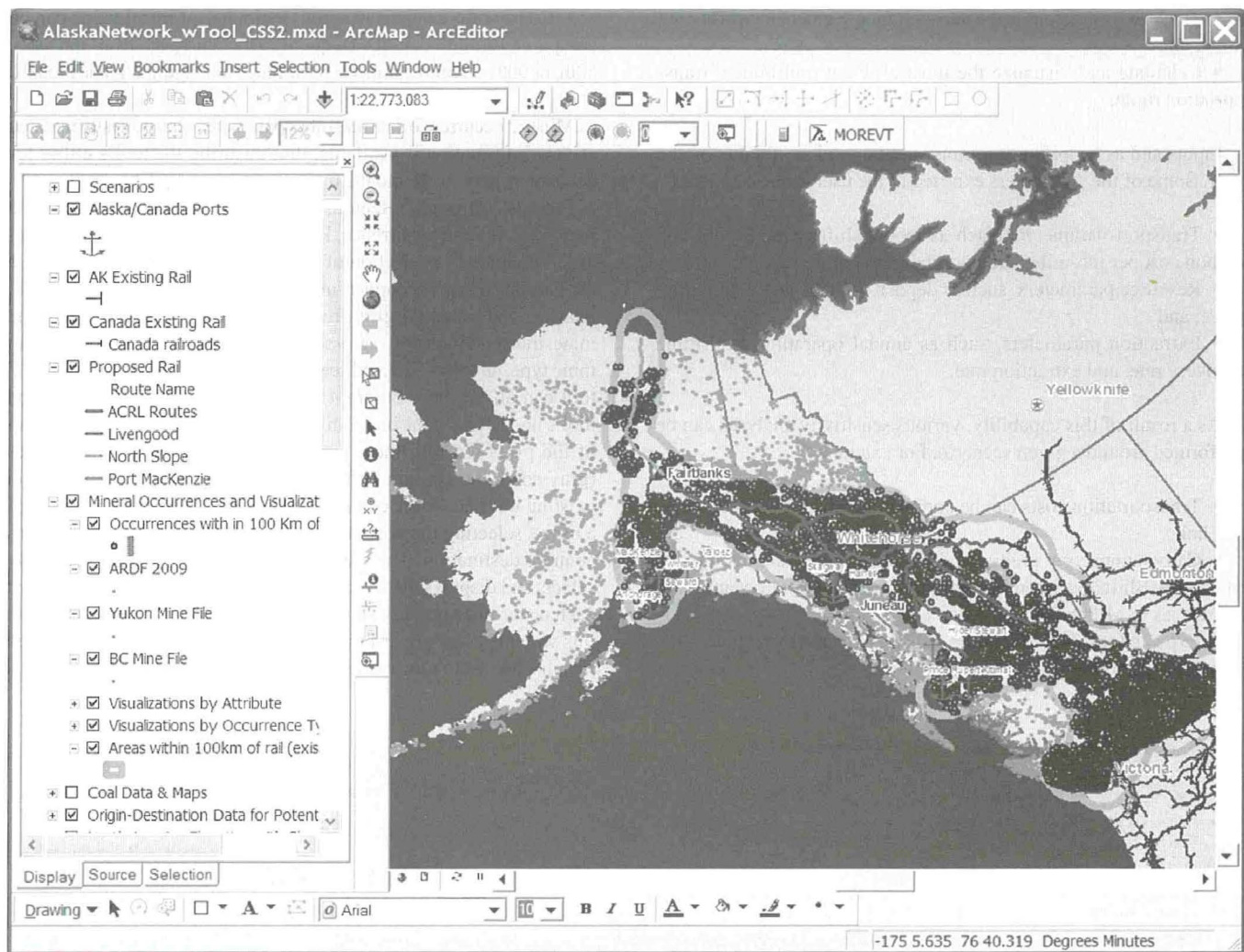


FIGURE 4 Example of ArcGIS interface displaying data that support mineral occurrence evaluation with MOREV.

For example, one early user discovered that most of the potential GMV in a custom corridor related to only two mineral deposit types. These would be areas for a developer to focus on in trying to ensure an economic rate of return on rail and mineral investments.

## DETAILED DESCRIPTION OF TOOL

### Data Sources

The data used in the tool consist of the marriage of mineral resource database files with relevant GIS layers from regional and national government sources, including existing and proposed rail transportation networks, augmented by expert knowledge. Metz has compiled a database of nearly 7,300 mineral occurrences in the region associated with a mineral deposit model based on work by Cox and Singer (4). This database has been joined with occurrence data from the Alaska Resource Data File (5) and similar data from Yukon Territories (6) and British Columbia (7). Additionally, the University of Alaska, Fairbanks, team has computed conservative probability of development factors for each occurrence, including mineral concentrate factors (ratio of molecular weight of ore to that of contained

metal), which allow the calculation of both metal weights for commodity valuation and weights of mineral concentrate for shipping calculations. For example, the parent shipped ore stibnite weighs 1.4 times as much as refined antimony. With this information, a developer could choose to ship mineral concentrate to a refiner rather than invest the capital costs for an on-site smelter.

Resource estimation involves predicting the tonnage and grade of a mineral deposit. Mineral deposit models capture the classification of mineral deposits according to their geologic makeup and economic value. They describe statistical spreads of ore grade and tonnage from similarly classed known mineral reserves. Based on knowledge of the frequencies of grades and tonnages that characterize similar deposits, these models (3) form the basis for quantifying predictive estimates. This study incorporates the intercepts representing the 10th, 50th, and 90th frequency of occurrence percentiles for the models; that is, choosing the 50th percentile model value for a particular occurrence gives the grade and tonnage of at least 50% of similarly classed deposits.

The transportation network data consist of the existing intermodal transportation network produced by Oak Ridge National Laboratory's Center for Transportation Analysis (8) integrated with the proposed ACRL as developed by the Phase I feasibility study (1). Integrating

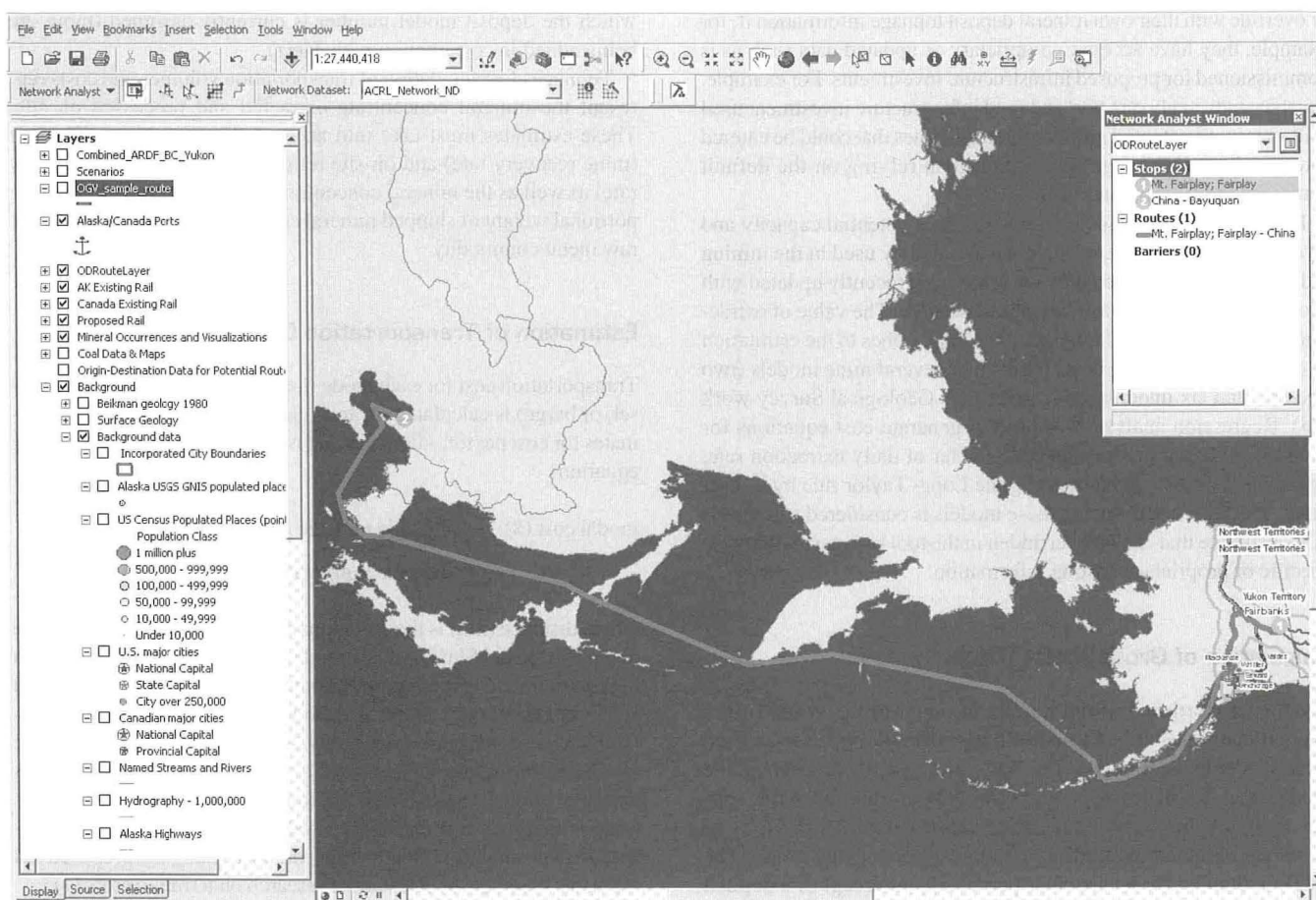


FIGURE 5 Example of transportation routing component of MOREV selected by user to calculate costs and fastest rail, road, and shipping routes to port of Bayuquan, China, for mineral concentrates extracted from Fairplay mineral occurrence.

these various data sources took careful investigation of publicly available data and recent scientific investigations (2, 3, 5–7, 9). However, using desktop ArcGIS and its flexible software development environment meant that potential integration challenges were surmountable through careful documentation of data sources. Additional documentation is also available through MOREV, which always provides traceback to original data sources via clickable help links.

## GIS Platform

The MOREV application has been developed using desktop ArcGIS (Version 9.3.1), the industry standard in GIS software technology created by the Environmental Systems Research Institute (ESRI). The MOREV development team used the ArcGIS development and design framework ArcObjects in conjunction with Microsoft's Visual Basic development platform to allow for the customization and extension of numerous GIS functions and capabilities intrinsic to the ArcGIS software. MOREV, by design, takes advantage of the customization and extension of GIS tasks, including transportation routing, resource mapping, geographic analysis, data reporting and compilation, and visualization, while integrating them into a single, easy-to-use tool for spatial analysis and evaluation of Alaskan mineral occurrences as well as existing and proposed transportation routes. Using ArcGIS and the ArcObjects development environment as the

design foundation for the MOREV application structure supports the project's plan to reach a large user base by using the ESRI ArcGIS server software platform for delivering MOREV functionality via the web.

Current work is under way to recode MOREV by using Microsoft's VB.Net software environment framework to port MOREV to the Internet. A subset of the functionality of the MOREV desktop application will be available that will be compatible and extendable with future versions of ESRI software, such as the newly released ArcGIS 10. This updated version, with additional detailed costing information, is anticipated to be available through cooperative agreements with the University of Alaska, Fairbanks–Michigan Tech development team, which holds rights to the model. The web version of MOREV, which will have the current simple costing model, is expected to be available for use in early 2011 at <http://www.mtri.org/mineraloccurrence.html>.

## Estimation of Mineral Resources

The mineral deposit models allow for statistical estimation of mineral concentrate grades and tonnages for the entire region under study. By default the tool selects the 50th percentile model intercept, meaning that the reported tonnages are seen in at least 50% of similarly classed mineral deposits. Users can also select 10th or 90th percentile values,

or override with their own mineral deposit tonnage information if, for example, they have access to proprietary or updated data such as is commissioned for proposed infrastructure investments. For example, a recent study analyzing port and road infrastructure investment used drill-measured potential gold resource estimates that could be entered into the deposit tonnage field rather than relying on the default statistical model estimates (10).

Empirical relationships between a mine's potential capacity and its ore tonnage have been developed and widely used in the mining industry (11). This established work has been recently updated with economic mine data from a larger data set (12). The value of extraction rate is the basis for factored model approaches to the estimation of capital and operating costs for each of several mine models (two open pit and six underground), after U.S. Geological Survey work (13). Regression analysis was used to generate cost equations for each model based on the single parameter of daily extraction rate, or capacity, as calculated by using the Long–Taylor rule by or user input. The cost derived using these models is considered a prefeasibility estimate that can be overridden in the tool by a user with more specific or proprietary costing information.

### Calculation of Gross Metal Value

GMV (net worth in current dollars of metal in the ground) of a deposit can be calculated on the basis of model or user-entered mineral deposit tonnages. For each commodity present in the model (e.g., for Model 21a, porphyry copper–molybdenum, copper, molybdenum, gold, and silver can be present), the tons of metal are obtained by multiplying deposit tonnage by grade. The GMV evaluated for a particular percentage intercept of a deposit model is then the tons of metal times the desired price per ton for each commodity:

$$GMV = \sum_{i=0}^n T_i G_i P_i$$

where

$n$  = number of commodities in the model;

$T_i$  = model-derived ore tonnage at either the 10th, 50th, or 90th percentile frequency intercept for the  $i$ th commodity;

$G_i$  = grade (range: 0 to 1) for the  $i$ th commodity; and

$P_i$  = price per metric ton of the  $i$ th commodity.

GMV is expressed in U.S. dollars for commodity prices (in U.S. dollars per metric ton) and expresses the total gross current value of metal in the ground, regardless of the actual amount of metal eventually extracted.

GMV is a valuable summary number for decision makers because it has been linked to positive economic impacts for the development region (14). For example, a 1998 study by Information Insights for the Fairbanks North Star Borough estimated the annual economic impact of the Fort Knox gold mine at \$104 million during the anticipated 12-year mine life; this amount was updated to \$181.7 million per year in 2004 (14). The estimated GMV at the time was \$1.2 billion. Together, these figures demonstrate how the economic impact of a mineral resource development can be equated to the sizeable GMV of a mineral deposit.

GMV for an entire region can be visualized geographically. Figure 6 shows an early instantiation of an interpolation data set for the 55% of the mineral occurrences in the 22,000-record database for

which the deposit model number is currently assigned (more are being added in consultation with Metz).

Tonnages for calculation of transportation volumes and costs represent the mineral concentrate extracted and processed on site. These estimates must take into account the extraction efficiency (mine recovery rate) and on-site refining efficiency (mill recovery rate) as well as the mineral concentrate factor, which gives the proportional weight of shipped mineral concentrate to the weight of the raw metal commodity.

### Estimation of Transportation Costs

Transportation cost for each mode (i.e., rail, truck, ocean-going vessel, or barge) is calculated on an aggregate basis by using fleet estimates for cost per ton-kilometer and is estimated using the following equation:

$$\begin{aligned} \text{modal cost (\$)} &= \text{shippable tonnage (mT)} \\ &\quad * \text{distance (km)} * \$/\text{mT-km} \end{aligned}$$

where distance (km) is the cumulative distance along a route for the given mode, and \$/mT-km is a mode-specific freight transport cost in current dollars per megaton-kilometer for which sourced default values are shown in Table 1. Research into these costs per megaton-kilometer revealed a wide range of valid values. The study uses these as a basis for reasonable default or starting values; users are always able to override these values if, for example, more appropriate or proprietary data are available. Future development will incorporate route-specific adjustments in freight transport costs along the proposed rail segments based on ACRL Phase I research on terrain factors (1).

A regression equation (as opposed to a single value) dependent on route distance is used to calculate ocean freight transportation cost. This equation is derived from an analysis of 2009 AXSMarine dry bulk shipping indices (Figure 7). Some routes had high variation (e.g., \$10/mT nadir versus \$26/mT peak). Rates varied seasonally with peaks (end of June and November) and troughs (early March and September). Rates vary with distance in a linear relationship ( $r^2 = .7213$ ) such that rate (\$/mT-km) =  $(-8.62 \cdot 10^{-8})x + 0.003216$ , where  $x$  = distance (km).

The total transportation cost estimate also includes the cost of modal shifts (e.g., from rail to ocean-going vessel at a port). Table 1 shows the default values in dollars per shippable tonnage, which are multiplied by total shippable tonnage to derive the total cost for each transition. Again, the user can override these values if more appropriate data are available.

### Optimal Routing Module

To provide users with more informative estimates of transportation costs, MOREV interfaces with ESRI's network analyst toolbox to calculate transportation routes using the modified intermodal transportation network. Figure 8 shows a screenshot of this tool. Routing is dynamically calculated from a user-defined mineral occurrence origin to a specified destination (ports, cities, or facilities in the United States, Canada, or overseas). The routing algorithm automatically selects the route that minimizes transportation cost according to the default cost values shown in the cost subform.

The user has the option of constraining the route through a particular node (port or city) if he or she has knowledge of legal or logistical obstacles that would necessitate a detour of the cost-efficient

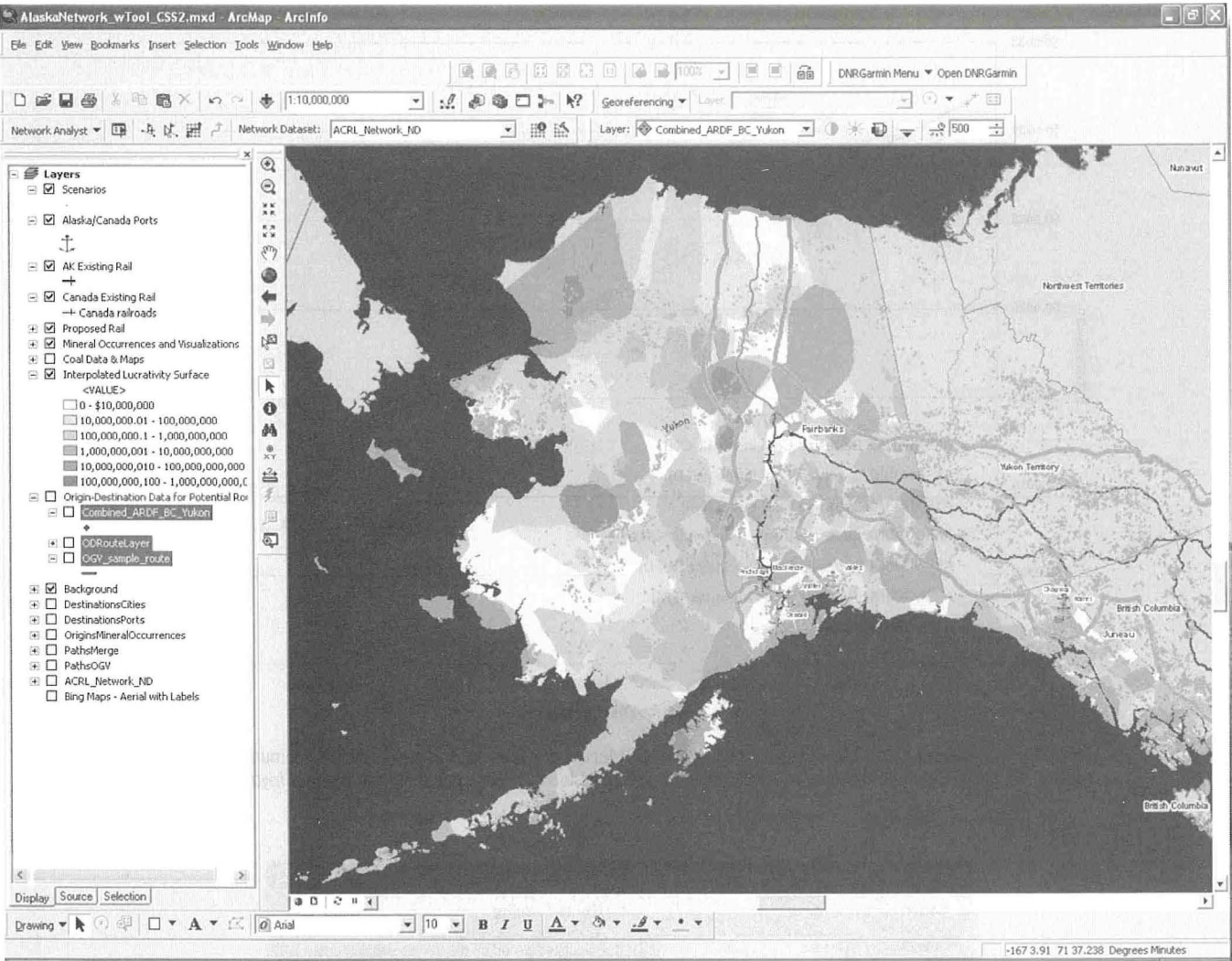


FIGURE 6 MOREV screen shot showing extrapolation through Kriging interpolation of GMV based on deposit model tonnage at 50th percentile for occurrences with valid deposit models in the region under study (light green = ~\$0; darkest green = \$950 billion).

TABLE 1 Default Freight Transportation Costs Used in MOREV Tool

	Source	Default Value
Mode		
Truck	Transportation Statistics Annual Report (15)	\$0.0940/mT-km
Rail	National Transportation Statistics (16)	\$0.0177/mT-km
Domestic water	National Transportation Statistics (16)	\$0.0320/mT-km
Ocean freight	AXSMarine dry bulk shipping indices (17)	$(-8.62 \cdot 10^{-8})x + 0.003216^a$
Intermodal transition		
Land–water	Mark Taylor (UAF) and publicly available port fee data	\$3.00/mT
Truck–rail	Bob Rieck, expert opinion (Agrico Sales)	\$4.50/mT

NOTE: UAF = University of Alaska, Fairbanks.  
“Where  $x$  = route distance (km).

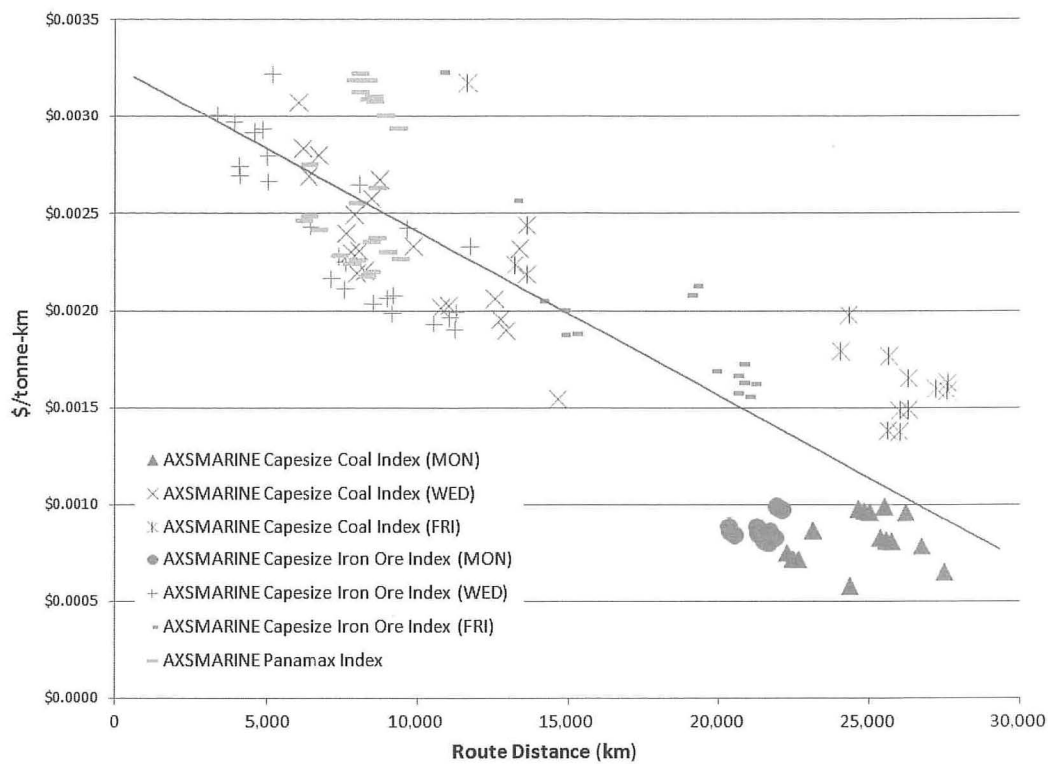


FIGURE 7 Shipping rates for 2009 according to AXSMarine dry bulk shipping indices for various oceanic routes. [Source: Shipping distances, AXSMarine distance table (17); cost per megaton rates, *Lloyd's List* (18).]

The screenshot shows the 'MOREVT -Origin/Destination and Routing Form' window. It is divided into several sections:
 

- Origin:** Region (Alaska), Mineral Occurrence (Mt. Fairplay/Fairplay), and a 'Load Origin from Selected' button.
- Destination:** Type (Port), Region (Alaska), and Name (Alaska - Anchorage).
- Output Options:** Checkboxes for 'Export KML of calculated route' and 'Export SHP of calculated route'.
- Route Forcing:** A section with a descriptive text and two identical blocks. Each block has a 'Force route through this location' checkbox (unchecked), followed by Type (Port), Region (Canada), and Name (Canada - Canaport) dropdowns.
- Buttons:** 'Calculate Route' and 'Close' at the bottom.
- Radio Buttons:** 'Use proposed rail' (selected) and 'Use just existing rail'.

FIGURE 8 MOREVT screen shot of optimal routing module form.

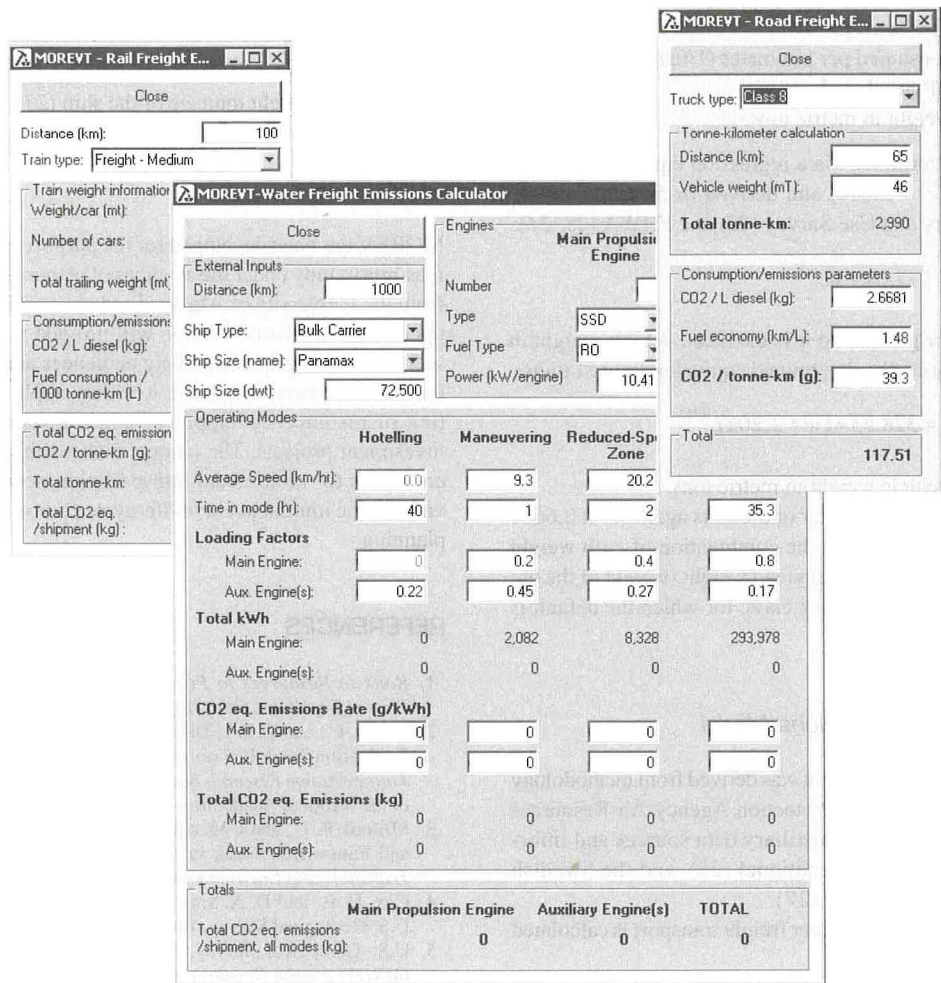


FIGURE 9 Mode-specific calculator forms for carbon emissions in transportation carbon accounting module.

route in real-world scenarios. The module also enables users to compare route options and costs with and without the proposed ACRL segments. After route calculation, distances for each transportation mode are transferred to the costing calculator.

Estimation of Carbon Impact

The transportation carbon accounting module calculates an estimate of CO<sub>2</sub>-equivalent emissions (which includes CO<sub>2</sub>, methane, and nitrous oxide) from the transportation distance and tonnage of shipped ore for a particular scenario. Unique emissions models have been researched and developed for road, rail, and water freight transportation that estimate carbon emissions rates per megaton-kilometer based on fuel usage.

Each mode-specific calculator form (Figure 9) makes the model assumptions transparent and user adjustable. The forms are prepopulated with reasonable default values, along with clickable explanations and the genealogy of the assumptions (19). Emissions rates are multiplied by shippable tonnage and mode-specific transportation distance to derive total emissions estimates. To address potential economic impacts, users are also able to assign a cost per megaton of carbon to include in total transportation cost estimates.

Rail Freight Carbon Emissions Model

The carbon emissions rate for rail freight transport is calculated as follows:

CO<sub>2</sub>/mT-km = F \* C

where *F* is the fuel consumption rate in liters per megaton-kilometer and *C* is CO<sub>2</sub> per liter of diesel. The default value for the fuel consumption rate is set at 0.005946 L/mT-km [converted to SI from the Association of American Railroads 2007 fleet average figure of 436 revenue-ton-miles per gallon of fuel consumed (20)]. The default value for CO<sub>2</sub> per liter of diesel is set at 2.6681, following the U.S. Environmental Protection Agency (EPA) (21). Research is under way to create a more dynamic rail carbon emissions model that links train weight to fuel economy.

Truck Freight Carbon Emissions Model

The carbon emissions rate for truck freight transport is calculated as follows:

CO<sub>2</sub>/mT-km = R \*  $\frac{C}{W}$

where

$R$  = liters of fuel consumed per kilometer (1/fuel economy),  
 $C$  = CO<sub>2</sub> per liter of diesel, and  
 $W$  = total vehicle weight in metric tons.

Fuel economy is determined via a regression equation based on total vehicle weight ( $r^2 = .9605$ ) and derived from data from the 2002 Vehicle Inventory and Use Survey (22) and FHWA (23, 24):

$$\text{fuel economy (mpg)} = 772.04 * w^{-0.463}$$

where mpg is miles per gallon and  $w$  is the mean vehicle weight in pounds. Fuel economy can be determined equivalently in SI units:

$$\text{fuel economy (km/L)} = 328.23 * (w * 2,204)^{-0.463}$$

where  $w$  is the mean vehicle weight in metric tons.

The default value for CO<sub>2</sub> per liter of diesel is again set at 2.6681 (21). Vehicle weight is defined as the combination of curb weight (weight of an empty vehicle) plus payload weight (weight of the ore being shipped) and is based on truck class, for which the default is "Class 8, combination" (22).

### Water Freight Carbon Emissions Model

The emissions model for water freight was derived from methodology of the California Environmental Protection Agency Air Resources Board (25) and augmented with auxiliary data sources and information from EPA (21), ICF International (26), and the Swedish Environmental Protection Agency (27).

The carbon emissions rate for water freight transport is calculated on a per trip basis:

$$\frac{\text{CO}_2}{\text{trip}} = \text{auxiliary engine emissions} + \text{main propulsion engine emissions}$$

where auxiliary engine emissions is the auxiliary engine activity in kilowatt hours times the auxiliary engine emissions factors, and main propulsion engine emissions is the propulsion engine activity in kilowatt hours times the propulsion engine emissions factors. Engine emissions factors in grams per kilowatt hour are the CO<sub>2</sub>-equivalent emissions measured by the Swedish Environmental Protection Agency (27).

Engine activity is calculated as follows:

$$\begin{aligned} \text{engine activity (kW-h)} &= \text{time-in-mode (h)} * \text{load factor (\%)} \\ &\quad * \text{engine power (kW-h) summed over} \\ &\quad \text{each activity mode (cruise, maneuver,} \\ &\quad \text{reduced speed zone, and berthing)} \end{aligned}$$

These variables are defined as follows:

- time-in-mode (h): defaults for cruise are (distance \* speed); defaults for other activity modes are from Thesing et al. (28);
- speed (km/h): dependent on ship type (29) and activity mode (26);

- load factor (%): dependent on ship type and engine type (21); and
- engine power (kW): determined by regression equation from the domestic weight tonnage of the ship (23).

### SUMMARY

MOREV has been developed for the purpose of extending an ACRL feasibility study under the auspices of resource development experts from the University of Alaska, Fairbanks. This desktop tool and its upcoming web-based version will provide a useful framework for scenario development for many audiences and users. Its extensibility and customizability make it widely applicable for joint evaluation of resource development and transportation infrastructure investment projects. The unique dual capability of estimating CO<sub>2</sub> emissions for multimodal transportation networks adds additional value in the imminent era of life-cycle carbon-budget-aware project planning.

### REFERENCES

1. *Rails to Resources to Ports: The Alaska Canada Rail Link Project Phase I Feasibility Study*. Executive report. ALCAN RailLink Inc., 2007.
2. Lai, Y.-C., and C. P. L. Barkan. Enhanced Parametric Railway Capacity Evaluation Tool. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2117, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 33–40.
3. Milford, R. L., and J. M. Allwood. Assessing the CO<sub>2</sub> Impact of Current and Future Rail Track in the UK. *Transportation Research Part D: Transport and Environment*, Vol. 15, No. 2, 2010, pp. 61–72.
4. Cox, D. P., and D. A. Singer. *Mineral Deposit Models*. Bulletin 1693. U.S. Geological Survey, 1986.
5. U.S. Geological Survey. *National Mineral Resource Data System (MRDS) Alaska Resource Data File (ARDF)*. <http://ardf.wr.usgs.gov>. Accessed March 16, 2009.
6. Deklerk, R. *Yukon MINFILE 2002: A Database of Mineral Occurrences*. Exploration and Geological Services Division, Yukon Region, Indian and Northern Affairs Canada, Whitehorse, Yukon, Canada, 2002.
7. Jones, L. *MINFILE Mineral Occurrence Database*. BCGOV EMPR British Columbia Geological Survey Branch, Victoria, British Columbia, Canada, 2006.
8. Peterson, B. *Intermodal Transportation Network*. Oak Ridge National Laboratory Center for Transportation Analysis, Oak Ridge, Tenn., 2009. [www-cta.ornl.gov/transnet/Intermodal\\_Network.html](http://www-cta.ornl.gov/transnet/Intermodal_Network.html). Accessed Aug. 30, 2009.
9. Zhang, K., R. Nair, H. S. Mahmassani, E. D. Miller-Hooks, V. C. Arcot, A. Kuo, J. Dong, and C.-C. Lu. Application and Validation of Dynamic Freight Simulation–Assignment Model to Large-Scale Intermodal Rail Network: Pan-European Case. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2066, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 9–20.
10. Lemke, K. A. Evaluation of River Port and Road Networks to Support Mineral Development in Western Alaska. Presented at 45th Annual Transportation Research Forum, Evanston, Ill., 2004.
11. Taylor, H. K. Mine Valuation and Feasibility Studies. In *Mineral Industry Costs*, 2nd ed. (J. R. Hoskins and W. R. Green, eds.), Northwest Mining Association, Spokane, Wash., 1978, pp. 1–17.
12. Long, K. R. A Test and Re-Estimation of Taylor's Empirical Capacity–Reserve Relationship. *Natural Resources Research*, Vol. 18, No. 1, 2009, pp. 57–63.
13. Camm, T. W. *Simplified Cost Models for Prefeasibility Mineral Evaluations*. Information Circular 9298. Bureau of Mines, U.S. Department of the Interior, 1991.
14. Liles, P. Fort Knox Seeks Mine Life Extension: Originally Due to Close This Year, Fort Knox Still Producing. *Alaska Business Monthly*, November 2005.

15. Research and Innovative Technology Administration, Bureau of Transportation Statistics, U.S. Department of Transportation. *Transportation Statistics Annual Report 2008*. [http://www.bts.gov/publications/transportation\\_statistics\\_annual\\_report/2008/](http://www.bts.gov/publications/transportation_statistics_annual_report/2008/). Accessed Aug. 3, 2009.
16. Research and Innovative Technology Administration, Bureau of Transportation Statistics, U.S. Department of Transportation. *National Transportation Statistics*. [http://www.bts.gov/publications/national\\_transportation\\_statistics/](http://www.bts.gov/publications/national_transportation_statistics/). Accessed Aug. 3, 2009.
17. AXSMarine. *AXSMarine Distance Table*. <http://www.axsmarine.com/distance/>. Accessed Feb. 20, 2010.
18. *Lloyd's List*. <http://www.lloydslist.com/ll/sector/dry-cargo/>. Accessed Feb. 18, 2010.
19. U.S. Environmental Protection Agency. *Emission Facts: Average Carbon Dioxide Emissions Resulting from Gasoline and Diesel Fuel*. <http://www.epa.gov/oms/climate/420f05001.htm>. Accessed Aug. 20, 2009.
20. *Railroad Facts 2008*. Association of American Railroads, Washington, D.C., 2008.
21. Energy and Environmental Analysis, Inc. *Analysis of Commercial Marine Vessels Emissions and Fuel Consumption Data*. Publication EPA420-R-00-002. Office of Transportation and Air Quality, U.S. Environmental Protection Agency, 2000.
22. U.S. Census Bureau. *Vehicle Inventory and Use Survey 2002*. <http://www.census.gov/svsd/www/vius/products.html>. Accessed Nov. 5, 2009.
23. FHWA, U.S. Department of Transportation. *Highway Statistics 2008*. <http://www.fhwa.dot.gov/policyinformation/statistics/>. Accessed Feb. 1, 2010.
24. Alam, M., and G. Rajamanickam. *Freight Analysis Framework, Technical Documentation: 2002 and 2035, Report Number S9: Development of Truck Payload Equivalent Factor*. FHWA, U.S. Department of Transportation, 2007.
25. Planning and Technical Support Division, Air Resources Board, California Environmental Protection Agency. *Statewide Marine Auxiliary Engine Emissions Inventory*. <http://www.arb.ca.gov/ports/marinevess/presentations.htm>. Accessed Aug. 1, 2009.
26. Browning, L. Current Methodologies and Best Practices for Preparing Port Emission Inventories. Presented at 15th International Emission Inventory Conference, New Orleans, La., 2006.
27. Cooper, D., and T. Gustafsson. *Methodology for Calculating Emissions from Ships. 1. Update of Emission Factors*. Swedish Methodology for Environmental Data, Swedish Meteorological and Hydrological Institute, Swedish Environmental Protection Agency, Norrköping, Sweden, 2004.
28. Thesing, K. B., A. Mannino, and A. Edwards. Nine Ports in the 49th State: Commercial Marine Inventory for Alaska. Presented at 15th International Emission Inventory Conference, New Orleans, La., 2006.
29. *Marine Exhaust Emission Quantification Study: Baltic Sea*. MEPC 45/INF.7. Lloyd's Register and International Maritime Organization, London, 1998.

---

*The Freight Transportation Economics and Regulation Committee peer-reviewed this paper.*

# Midwest Rail Study

## Modeled Near-Field Impacts of Emissions of Fine Particulate Matter from Railyard Activities

Stephen N. Feinberg, Varun Yadav, Jeremy G. Heiken, and Jay R. Turner

A detailed inventory of emissions including fine particulate matter ( $PM_{2.5}$ ) was developed for the CSX Rougemere railyard in Dearborn, Michigan, for 2007 to 2008. Emissions were temporally allocated at monthly, day-of-week, and hourly levels and spatially allocated within the railyard footprint. Dispersion modeling was conducted with AERMOD to estimate railyard near-field impacts including concentration levels at the Dearborn compliance monitoring station located about 150 m from the railyard fenceline. Annual average impacts ranged from 0.3 to 0.5  $\mu\text{g}/\text{m}^3$  in 2007 and 0.2 to 0.3  $\mu\text{g}/\text{m}^3$  in 2008, depending on the spatial allocation of locomotive activities within the railyard. The reduction in railyard impacts was caused primarily by the reduction of particulate matter emissions that occurred after conventional switcher locomotives were replaced with locomotives with the new GenSet locomotive technology. The proximity of the railyard to the Dearborn monitor and the dominant role of switcher locomotive emissions motivated an examination of the spatial allocation of the switcher locomotives on modeled impacts. Monte Carlo simulations were performed to randomly place switcher locomotive activity on specific tracks within the railyard. This approach yielded high variability for modeled hourly impacts at the Dearborn monitor. However, the variability was relatively small for the modeled daily average impacts and even smaller for the modeled annual average impacts.

Transportation hubs such as ship ports, airports, and railyards can be localized sources of off-road mobile source air pollutant emissions. The footprints for such hubs are often located near communities, and air quality impacts may be a concern. Recent interest in fine particulate matter ( $PM_{2.5}$ ) hotspots has led to assessments of the impact of transportation hub emissions on near-field air pollutant burdens. For the case of railyards, emissions from diesel sources are typically the primary concern. For example, long-term monitoring of ambient PM concentrations has been conducted near a large railyard in Roseville, California (1). The measurements included  $PM_{2.5}$  mass, elemental carbon, and black carbon, with the latter two parameters being a surrogate for diesel PM emissions. For the summer 2007 monitoring season the average nighttime (2200 to 0500 PST)  $PM_{2.5}$  mass concentration difference across the railyard was  $\sim 8 \mu\text{g}/\text{m}^3$  for

the subset of data meeting the upwind-downwind screening criteria. Most of the mass difference was attributed to diesel PM emissions within the railyard.

The examination of the Roseville facility was just one component of a broader quantification of the air quality impacts of 17 California railyards on local ambient air pollutant concentrations. From 2004 to 2009, 17 distinct health risk assessment studies were completed for individual railyards or groupings of proximate railyards in California (2). These assessments covered a range of railyard types including classification, intermodal, maintenance, and specialty yards. The characterization of emissions and air quality impacts varied significantly by yard type, in which the proportion of total facility  $PM_{2.5}$  emissions originating from locomotive sources ranged from about 25% to over 95% depending on the types of operations present (3).

Other recent attempts have been made to better model locomotive emissions and their air quality impacts (4, 5). A study performed on the Alameda Corridor in California modeled the effects of multiple railyards and line haul operations around the corridor. The study focused on  $PM_{10}$  and oxides of nitrogen ( $NO_x$ ) emissions and their impacts within the corridor, where their estimated maximum 24-h contributions were 2.1 and 75  $\mu\text{g}/\text{m}^3$ , respectively (6). However, to better assess near-field impacts, refined modeling methods may be needed to account for the spatial distribution of emissions within a railyard or intermodal center. For example, Corfa et al. (7) modeled the impacts of frequent and worst-case emissions scenarios at a bus station and a railway station using computational fluid dynamics software.

The analyses presented in this paper were performed for the Midwest Rail Study, which is being conducted by the U.S. Environmental Protection Agency (EPA) Region V to quantify  $PM_{2.5}$  emissions from railyards and their impacts on proximate ambient air quality monitoring sites. Phase I included the development of an emission inventory for the CSX Rougemere railyard in Dearborn, Michigan; dispersion modeling of railyard-related emissions activities; and 3 months of hourly carbonaceous  $PM_{2.5}$  measurements at three sites in Dearborn (8). Two monitoring sites were within 150 m of the railyard fenceline, and the remaining monitoring site was 6 km from the railyard. While the Rougemere railyard is a relatively small classification, one of the proximate monitoring sites, and the focus of this analysis, was the Dearborn air quality compliance monitoring station operated by the Michigan Department of Environmental Quality (Michigan DEQ). The site is located on the footprint of a school and is adjacent to a residential neighborhood. Historically, the highest annual average and daily average  $PM_{2.5}$  mass concentrations in Michigan have been recorded at the Dearborn station, which is within the seven-county southeast Michigan fine PM nonattainment area encompassing metropolitan Detroit. Only recently has this nonattainment area met the national ambient air quality standards

---

S. N. Feinberg, V. Yadav, and J. R. Turner, Department of Energy, Environmental, and Chemical Engineering, Washington University in Saint Louis, Campus Box 1180, 1 Brookings Drive, Saint Louis, MO 63130-4899. J. G. Heiken, Sierra Research, Inc., 1801 J Street, Sacramento, CA 95811. Corresponding author: J. R. Turner, jrturner@wustl.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 106–114.  
DOI: 10.3141/2261-12

(NAAQS) for both annual and 24-h average  $PM_{2.5}$ . Previous analyses have demonstrated significant contributions from local emission sources to the  $PM_{2.5}$  mass concentration levels observed at the Dearborn station (9), which has raised the importance of quantifying impacts from the railyard and other nearby emissions sources, including a steelworks. The Rougemere railyard was also singled out by Michigan DEQ for evaluation and control in the recent Michigan State Implementation Plan because of its potential influence on the Dearborn monitor (10). During 2007 to 2008, the period covered by the emission inventory and dispersion modeling presented in this paper, the four switcher locomotives at the CSX Rougemere railyard were replaced with low-emitting GenSet technology locomotives manufactured by the National Railway Equipment Company.

## METHODS, MATERIALS, AND DATA

### Inventory of Emissions

Minimal federal guidance exists on procedures for the preparation of emission inventories for railyards. Moreover, the variability in activities from one yard to the next prohibits the development of standard guidelines that would be accurate in any given situation. In short, the best procedures for inventory development are from the ground up using facility-specific data whenever possible.

As part of the Midwest Rail Study, a detailed facility-specific inventory of total hydrocarbons, carbon monoxide,  $NO_x$ , sulfur dioxide, and  $PM_{2.5}$  emissions was developed for the Rougemere railyard. This inventory was made possible by the exceptional assistance provided by the CSX Corporation (the yard's owner) and its employees. The inventoried sources were comprehensive and included locomotives, nonroad equipment, on-road vehicles, and stationary sources. Details of the Rougemere emission inventory methods are documented by Heiken (11), who adopted previous efforts (12). The inventory methods and results are briefly summarized here.

Emissions-generating activities at Rougemere were classified into two major groups: locomotives and nonlocomotives. The specific activities included in the locomotive and nonlocomotive groups are as follows (11):

- Locomotives:
  - Through locomotives,
  - Arrival–departure locomotives,
  - Additional idling from arrival–departure locomotives,
  - Switcher locomotive operation, and
  - Switcher refueling idling and
- Nonlocomotives:
  - Worker vehicle exhaust,
  - Worker vehicle evaporative,
  - Heavy-duty diesel truck delivery,
  - Heavy-duty diesel truck delivery idle,
  - Facility truck,
  - Facility truck idle,
  - Space heating,
  - Water heating,
  - Liquid propane gas (welders, cutters),
  - Diesel (specialty vehicle carts),
  - Diesel (rubber tire loaders),
  - Diesel (forklifts),
  - Diesel (other general industrial equipment),
  - Diesel (snowblowers), and
  - Aerosol paints.

Emissions from locomotive auxiliary power units when present and operating were counted as part of the locomotive emissions total. Locomotive emissions were calculated from emissions factor data and yard-specific activity estimates. Notch-specific locomotive emissions factors were obtained by Sierra Research over multiple studies including data provided by the National Railway Equipment Company specifically for the GenSet technology switcher locomotive. Nonlocomotive emissions were calculated from emissions factors from standard EPA references (i.e., the MOBILE6.2 and NONROAD2005 models and AP-42) and yard-specific activity data. Yard-specific activity data were based on interviews of railyard personnel, train schedules, employee and switcher shift assignments, yard clock data, on-site observations, and other available emissions resources provided by CSX. Additional inventory inputs included local ambient conditions (e.g., temperature) and fuel parameters (e.g., diesel sulfur content).

The emission inventory was temporally allocated by month, day of week, and hour of day. The exception to this was road train events, which could be not resolved accurately at the hourly level because of network loading and other logistical difficulties; thus, road train activity was assigned uniform activity throughout the day. Overall total inventories for all sources were aggregated into daily totals (reported in units of pounds per day) and estimated for each date in the study period (January 1, 2007, through December 31, 2008). Annual emissions were then estimated by the sum of the days in the year. Hourly emissions estimates were renormalized into hourly diurnal profiles so that daily total emissions could be reassigned into hourly values as needed.

Figure 1 shows annual  $PM_{2.5}$  emissions for each locomotive activity and the aggregate of all nonlocomotive activities. Annual  $PM_{2.5}$  emissions decreased from 1.5 tons/year in 2007 to 1.0 ton/year in 2008, with most of these reductions realized from the phased replacement of all four railyard switching locomotives with GenSet technology locomotives (the first two locomotives were replaced in March 2008, and the remaining two locomotives were replaced in August 2008). Nonlocomotive activities were 7% and 10% of the 2007 and 2008 annual emissions, respectively. Year 2008 nonlocomotive emissions were 60% from diesel engines and 38% from space heating. Overall, the single greatest source of PM emissions at Rougemere is from switcher locomotive operation, even after the switchover to GenSet technology.

Figure 2 shows the spatial allocation of locomotive activities. The Michigan DEQ Dearborn monitoring station is also shown. Emissions are nonuniformly distributed over the railyard footprint. Furthermore, activity-specific actual emissions during any given hour will not be uniformly distributed over the regions highlighted in Figure 2. This presents one of the major challenges to hourly estimation of the railyard emissions impacts that was desired for comparison to hourly ambient  $PM_{2.5}$  carbon concentration data collected as part of the Midwest Rail Study: the source-to-receptor distances are on the same scale as the railyard footprint, and thus the location of activities within the railyard can profoundly influence the concentration levels at the receptor (monitoring) sites.

### Dispersion Modeling

Dispersion modeling was performed to estimate railyard activity impacts with emphasis on predicted  $PM_{2.5}$  mass concentrations at the Dearborn station. Nonlocomotive activities were not included in the dispersion modeling because they account for at most 10% of total

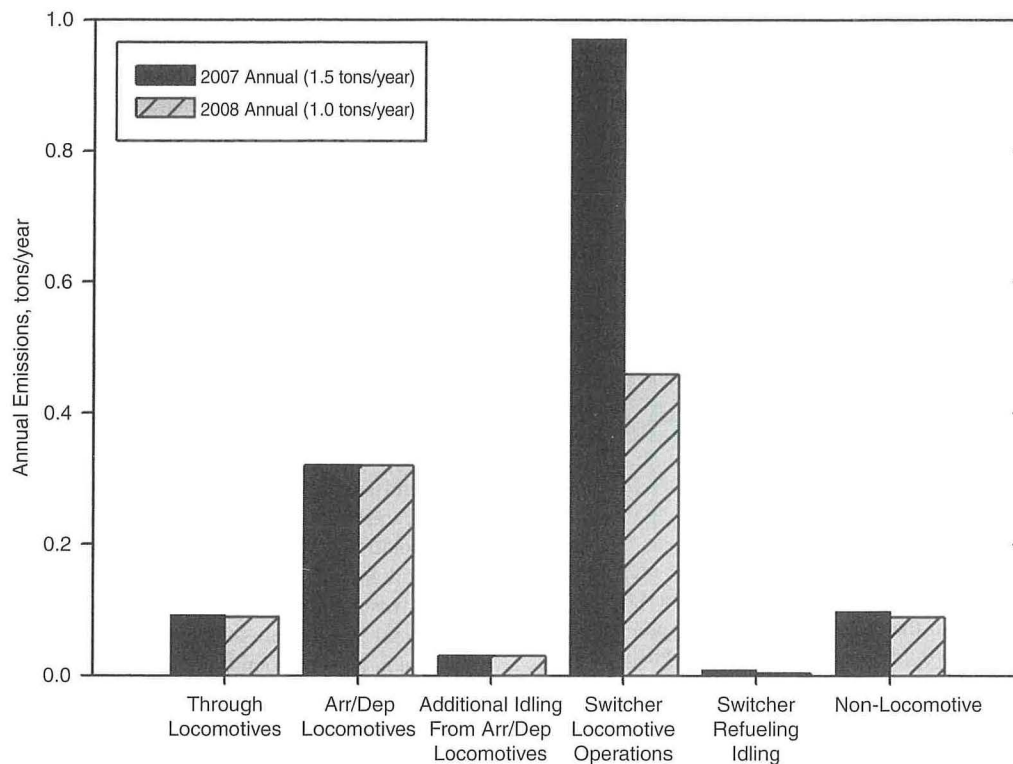


FIGURE 1 Annual emissions estimated from inventory of emissions at CSX Rougemere railyard (arr/dep = arrival-departure) (11).

PM<sub>2.5</sub> emissions. All locomotive operations were modeled as area sources representing a specific track segment with a 1.435-m width (based on the standard gauge used for tracks in the United States) and emission heights of 5.2 m for switchers and 6.1 m for road locomotives (this term is used to represent through locomotives and arrival-departure locomotives). Through tracks were divided into six regions of roughly equal length, each with two segments to represent the two parallel tracks. Switcher tracks were divided into 62 segments representing the ladder of parallel tracks (where trains are broken down and built up) and additional tracks to the southeast. Switcher refueling tracks included two segments corresponding to the labeled tracks in Figure 2. Coordinates for each of these area sources were determined using Google Earth.

With these track segments and the hourly specific emission inventory, a base case and three additional scenarios were modeled.

For the base case,

- Switcher locomotive operations emissions were homogeneously distributed over the switcher track segments;
- Through locomotives emissions were homogeneously distributed over the through track segments;
- Arrival-departure locomotives emissions were allocated with 80% to the through track segment directly across from the railyard operations office and 20% allocated to the remaining five through track segments;
- Additional idling from arrival-departure locomotives emissions was allocated 100% to the through track segment directly across from the railyard operations office; and
- Switcher refueling idling emissions were homogeneously distributed over the two switcher refueling track segments.

Scenario A was the same as the base case except arrival-departure locomotives emissions were allocated homogeneously across the through tracks (rather than allocating 80% of the emissions to the through track segment directly across from the railyard operations office and 20% to the remaining five through track segments). Scenario B was the same as the base case except 60% of the switcher locomotive operations emissions were allocated to the outermost ladder track (the ladder closest to the Dearborn station), and 40% were homogeneously allocated to the remaining switcher track segments (rather than all emissions being homogeneously distributed over the switcher track segments). This scenario is consistent with the observation that switcher locomotives tend to idle on the outermost ladder track (Track 29) between assignments.

Modeling with AERMOD was performed using flat terrain. AERMOD-ready meteorology data files were obtained from Michigan DEQ that were generated using AERMET with surface meteorology from the Detroit Metropolitan Wayne County Airport and upper air meteorology from White Lake, Michigan (World Meteorological Organization No. 72632). A 4.5-m-high receptor was placed at the Dearborn station. Modeling was also performed with grids of nine receptors arranged in a 3 × 3 array with 20-m spacing with the center receptor at the monitoring station. The array was oriented with faces perpendicular to the four cardinal directions. For certain time periods, the spatial distribution of railyard impacts was also examined by generating PM<sub>2.5</sub> mass concentration contour maps for AERMOD-modeled impacts with spatial emissions allocation by Scenario B. An 81 × 81 receptor grid was used with 6.25-m spacing between each receptor.

The sensitivity of modeled receptor concentrations to switcher activity within the railyard was examined by performing Monte Carlo

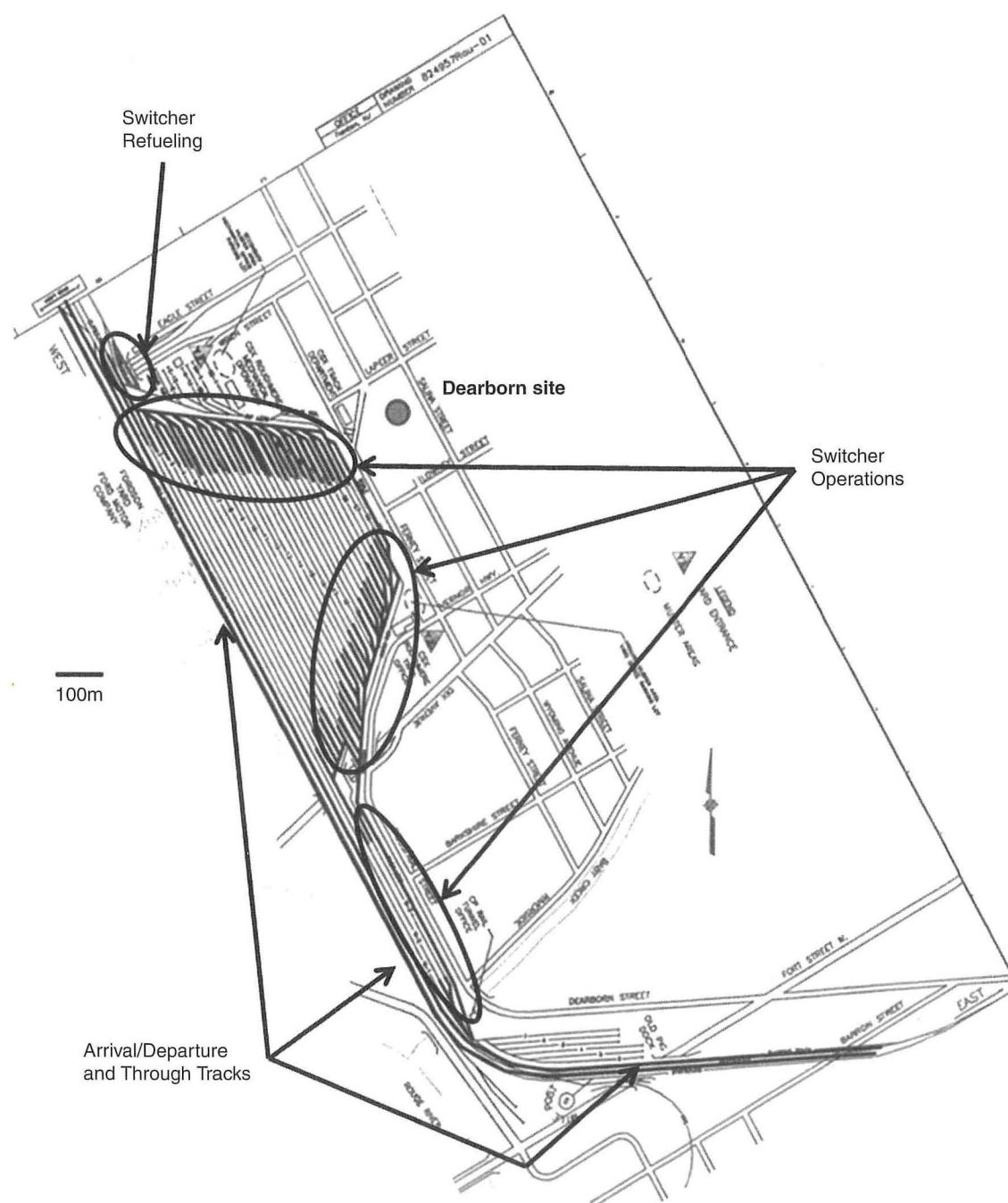


FIGURE 2 Spatial zones for arrival-departure locomotives, switcher operations, and switcher refueling. [Source: Adapted from Heiken (11).]

simulations. For each hour, switcher emissions were randomly assigned to only four of the track segments (because there are four switcher locomotives), and the concentration fields were modeled. This process was repeated 100 times for each hour to obtain a distribution of 100 predicted concentrations for that hour. While a switcher likely does not stay on a single switcher track segment for an entire hour, it is presumed that it operates in sufficient proximity to a single segment to yield similar impacts at the receptor.

## RESULTS

### Annual Average Impacts at the Dearborn Station

Dispersion modeling of the railyard locomotive emissions was performed for each hour over the years 2007 through 2008, consistent with the time period for the emission inventory. Of the total hours, 11% were not modeled because of calm winds. Modeled annual average concentrations are summarized in Table 1. For the base case, the modeled annual average  $PM_{2.5}$  mass concentrations at the Dearborn station from all locomotive activities were  $0.3 \mu\text{g}/\text{m}^3$  in 2007 and  $0.2 \mu\text{g}/\text{m}^3$  in 2008. The following patterns were observed for the Dearborn station: (a) spatial redistribution of the arrival-departure locomotives emissions (Scenario A) had negligible effect on the modeled annual average concentration; (b) spatially weighting the switcher locomotive operations emissions to the tracks nearest the Dearborn station (Scenario B) increased the modeled annual average concentration by ~60%; and (c) switcher locomotive operations emissions dominated the locomotive contributions to the modeled concentration, accounting for 83% to 90% of the 2007 annual average impacts and 71% to 82% of the 2008 annual average impacts. Modeled 2008 annual average impacts at the Dearborn station are 1% (base case) to 2% (Scenario B) of the annual average  $PM_{2.5}$  NAAQS of  $15 \mu\text{g}/\text{m}^3$ , which suggests the railyard is a relatively small contributor for this standard. However, all of the dispersion modeling results neglect hours with calm winds.

TABLE 1 AERMOD-Modeled Annual Average Receptor Concentrations of Emissions from Locomotive Operations at CSX Rougemere Railyard

	Modeling Scenario		
	Base	A	B
Annual average $PM_{2.5}$ mass ( $\mu\text{g}/\text{m}^3$ )			
2007	0.32	0.33	0.50
2008	0.19	0.19	0.30
Contributions by activity, 2008 (%)			
Through locomotives	4	4	3
Arrival-departure locomotives	11	12	7
Arrival-departure additional idle	1	1	<1
Switcher operations	84	83	90
Switcher refueling	<1	<1	<1

### Daily Average Impacts at Dearborn Station

Figure 3a shows the cumulative distributions of the year 2008 daily average impacts at the Dearborn station for base case emissions. Daily averages were calculated if there were at least 18 noncalm hours in the day; 39 days failed to meet this criterion. Maximum and 90th percentile daily average impacts were  $1.7$  and  $0.6 \mu\text{g}/\text{m}^3$ , respectively, which suggests the railyard is a relatively small contributor to the daily average  $PM_{2.5}$  NAAQS of  $35 \mu\text{g}/\text{m}^3$  (although all dispersion modeling results neglect hours with calm winds).

### Hourly Average Impacts at Dearborn Station

While there is no hourly  $PM_{2.5}$  NAAQS, dispersion modeling results at the hourly level were examined for comparison with the  $PM_{2.5}$  carbonaceous aerosol measurements collected as part of the Midwest Rail Study (8). Figure 3b shows the cumulative distributions of the

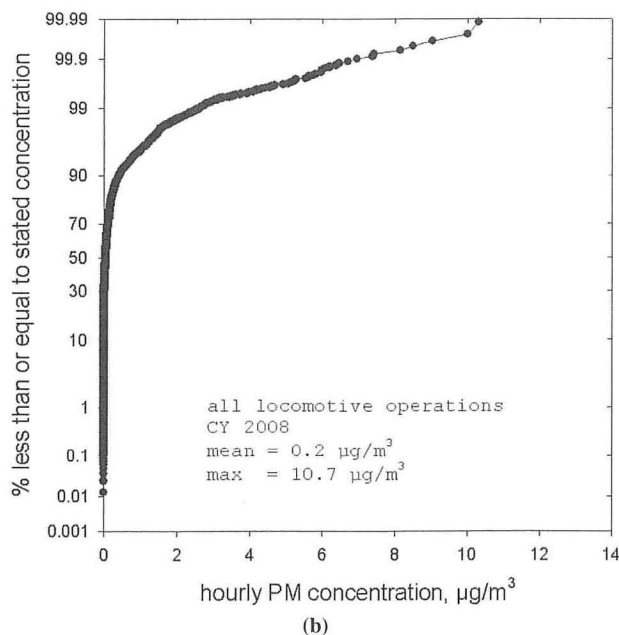
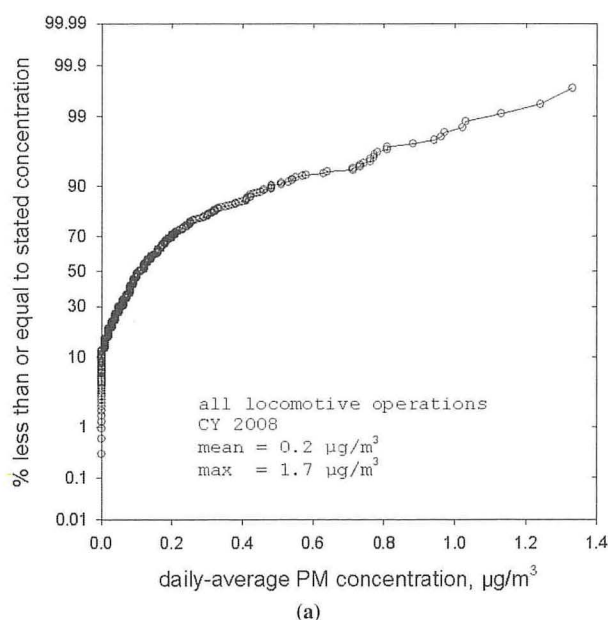
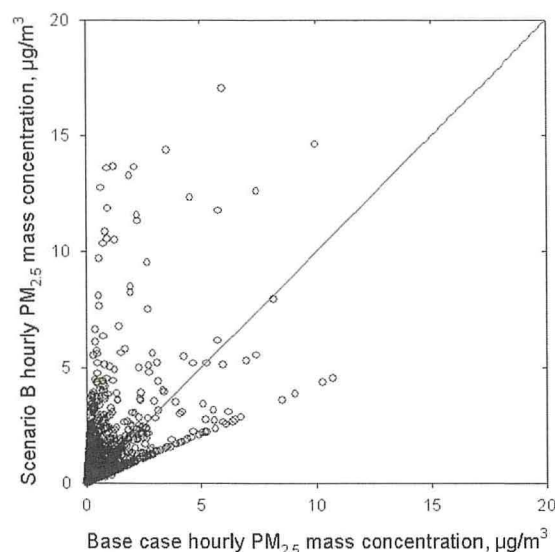


FIGURE 3 Cumulative modeled concentration distributions for locomotive impacts at Dearborn station (2008) for (a) daily and (b) hourly average  $PM_{2.5}$  (CY = calendar year).



**FIGURE 4** Modeled hourly concentrations at Dearborn receptor from railyard locomotives (2008) for base case and Scenario B assumptions for spatial allocation of switcher locomotive operations.

year 2008 hourly average impacts at the Dearborn station for base case emissions (note that the x-axis scale differs from that in Figure 3a by a factor of 10). Maximum and 90th percentile hourly average impacts were 11 and  $0.4 \mu\text{g}/\text{m}^3$ , respectively; 82% of hours had a modeled concentration less than  $0.2 \mu\text{g}/\text{m}^3$ .

The sensitivity of modeled impacts to the spatial allocation of switcher locomotive operations activities within the railyard is demonstrated in Figure 4, which is a scatter plot of hourly concentrations at the Dearborn station for Scenario B versus the base case. Dramatically different impacts are modeled depending on the spatial allocation of switcher locomotive emissions. This observation motivated using Monte Carlo simulations to better understand the distribution of impacts that could result depending on the switcher locomotives' locations within the railyard. Figure 5 shows the modeled concentration distributions obtained for 5 h that were chosen because they represent advective winds ( $2.6 \text{ m/s}$ ) from directions  $190^\circ\text{N}$  to  $270^\circ\text{N}$  at  $20^\circ$  intervals. For this analysis, 1,000 simulations were performed for each of the 5 h to stabilize the distributions, and the modeled concentrations were normalized by the respective concentrations obtained using the base case assumption of switcher emissions being uniformly distributed over the switcher tracks (Figure 2). While the base case modeled concentrations are significantly different across these 5 h (range,  $0.02$  to  $1.3 \mu\text{g}/\text{m}^3$ ), the normalized frequency distributions are quite similar. For four of the five modeled hours, the concentration distributions have a mode at 0% to 25% of the base case concentration. This corresponds to the assignment of the switchers to track segments within the switcher activity zone that at most weakly affect the Dearborn station. Figure 5d shows the mode is at 25% to 50% of the base case concentration because there are significant contributions from arrival-departure locomotives with 80% of their emissions allocated to the through tracks directly across from the railyard operations office. Maximum modeled concentrations can be as high as six times the base case modeled concentration. Such occurrences are rare and correspond to multiple switchers being randomly assigned to track segments that are directly upwind of the Dearborn station.

Figure 6a shows the median (open circles) and interquartile range (75th and 25th percentiles, error bars) of modeled hourly average  $\text{PM}_{2.5}$  mass concentrations at the Dearborn receptor for base case modeled median hourly impacts greater than  $2 \mu\text{g}/\text{m}^3$ . The modeled distributions for each hour were generated using 100 simulations, which yielded stable estimates for the median and interquartile range; extreme percentiles, such as the 5th and 95th percentiles, are likely to be unstable. Virtually all of the median values are less than the corresponding base case concentration, which means that hour-specific impacts are likely to be less than those predicted by the base case. Moreover, the interquartile ranges demonstrate that for each hour there is a very broad range of realizable impacts at the receptor depending on the precise location of the switcher locomotive operations within the railyard. In the context of the Midwest Rail Study, with hourly measurement of ambient PM parameters, real-time surveillance of switcher locomotive locations within the railyard would be needed to facilitate a direct comparison of modeled railyard impacts and observed concentrations at such time scales. Figure 6b shows the Monte Carlo-generated medians and interquartile ranges for daily average concentrations with median impacts greater than  $0.2 \mu\text{g}/\text{m}^3$ . In contrast to the hourly results, the interquartile ranges for the daily average impacts are much narrower, and thus the base case modeling provides a reasonable estimate of railyard impacts in the context of 24-h average  $\text{PM}_{2.5}$  NAAQS.

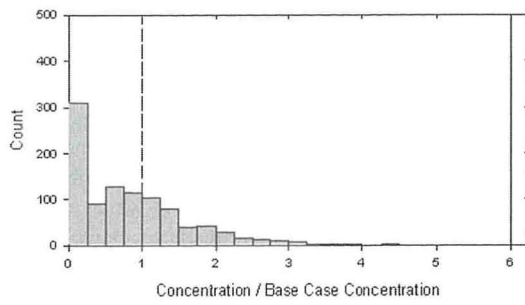
### Spatial Distribution of Daily Average Impacts

Figure 7a shows the modeled spatial distribution of daily average  $\text{PM}_{2.5}$  contributions from railyard emissions over a  $1\text{-km}^2$  area for November 19, 2007. This date represents a modeled 95th percentile impact day at the Dearborn receptor during September to December 2007, before the implementation of the GenSet switcher locomotives. The largest railyard  $\text{PM}_{2.5}$  contributions occur over the track closest to the Dearborn monitor, where the majority of switcher locomotive idling takes place. Twenty-four-hour average concentrations exceeding  $2 \mu\text{g}/\text{m}^3$  are predicted over a large portion of the school grounds.

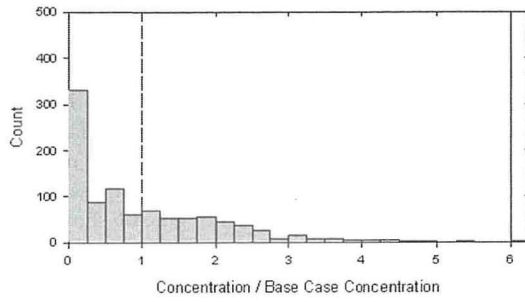
Figure 7b shows the modeled concentration field using November 17, 2008, emissions and November 19, 2007, meteorology. These dates correspond to the same day of the week (Monday) and were chosen to capture the changeover of all four switcher locomotives to the GenSet technology. Modeled contributions from the railyard dramatically decreased after the switchers were replaced. The 24-h average maximum concentration observed within the railyard footprint decreased from  $\sim 14$  to  $\sim 3 \mu\text{g}/\text{m}^3$  after switcher replacement. Modeled railyard impacts over the majority of the school grounds, including the Dearborn monitor, were  $0.4$  to  $0.9 \mu\text{g}/\text{m}^3$ .

### CONCLUSIONS

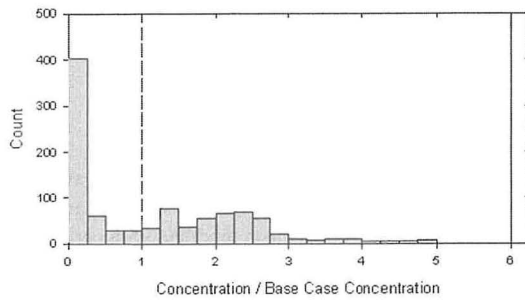
An emission inventory developed for the CSX Rougemere railyard in Dearborn showed a significant reduction in  $\text{PM}_{2.5}$  emissions from 2007 to 2008. This is largely because of the transition from conventional switcher locomotives to GenSet technology and additional reductions in diesel fuel sulfur content. These emissions reductions are reflected in the dispersion modeling, which estimates lower railyard emissions impacts at the Dearborn compliance monitoring station, located about 150 m from the railyard fenceline, in 2008 compared with 2007. The Rougemere facility is a relatively small-classification railyard, and the modeled  $\text{PM}_{2.5}$  mass concentrations



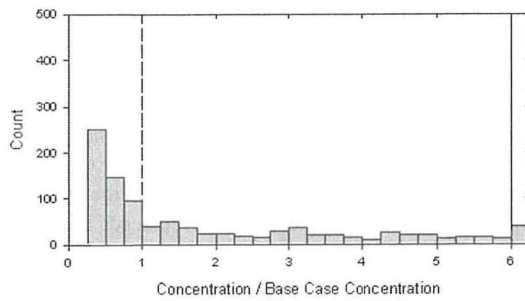
(a)



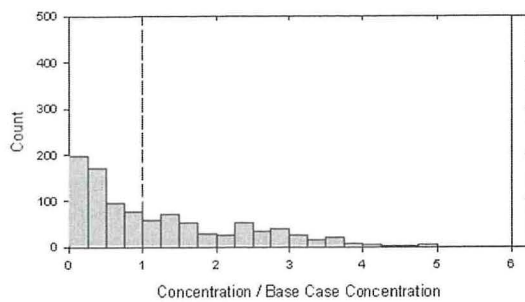
(b)



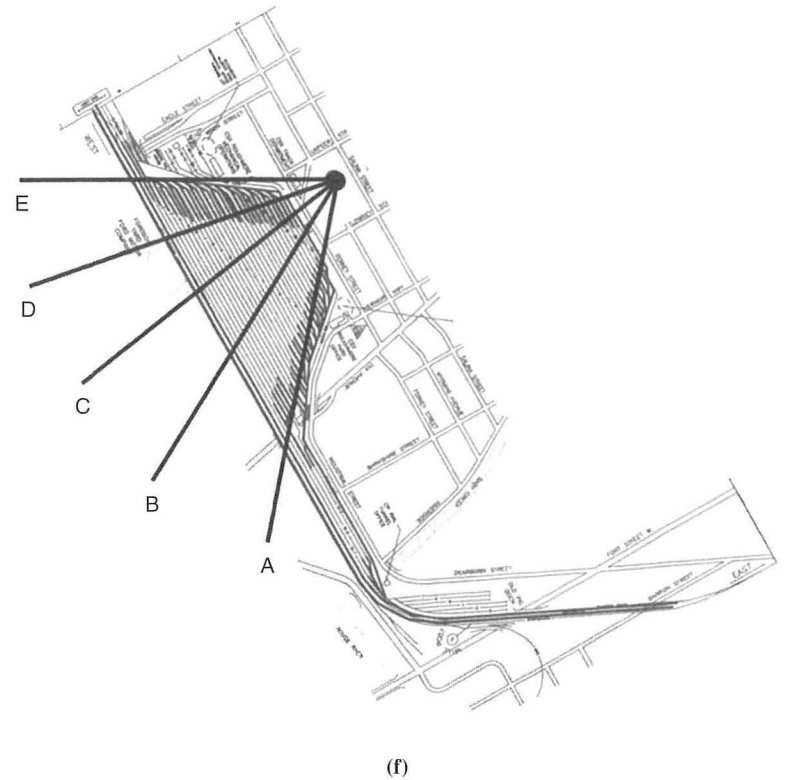
(c)



(d)



(e)



(f)

FIGURE 5 Frequency distributions of modeled hourly impacts for locomotives at Dearborn station from Monte Carlo simulations that randomly placed four switcher locomotives within their activity zones. Histograms: (a) E, (b) D, (c) C, (d) B, and (e) A [bars to right of solid vertical line (at 6 on x-axis) are numbers of simulations with scaled concentration  $\geq 6$ ]; (f) zones.

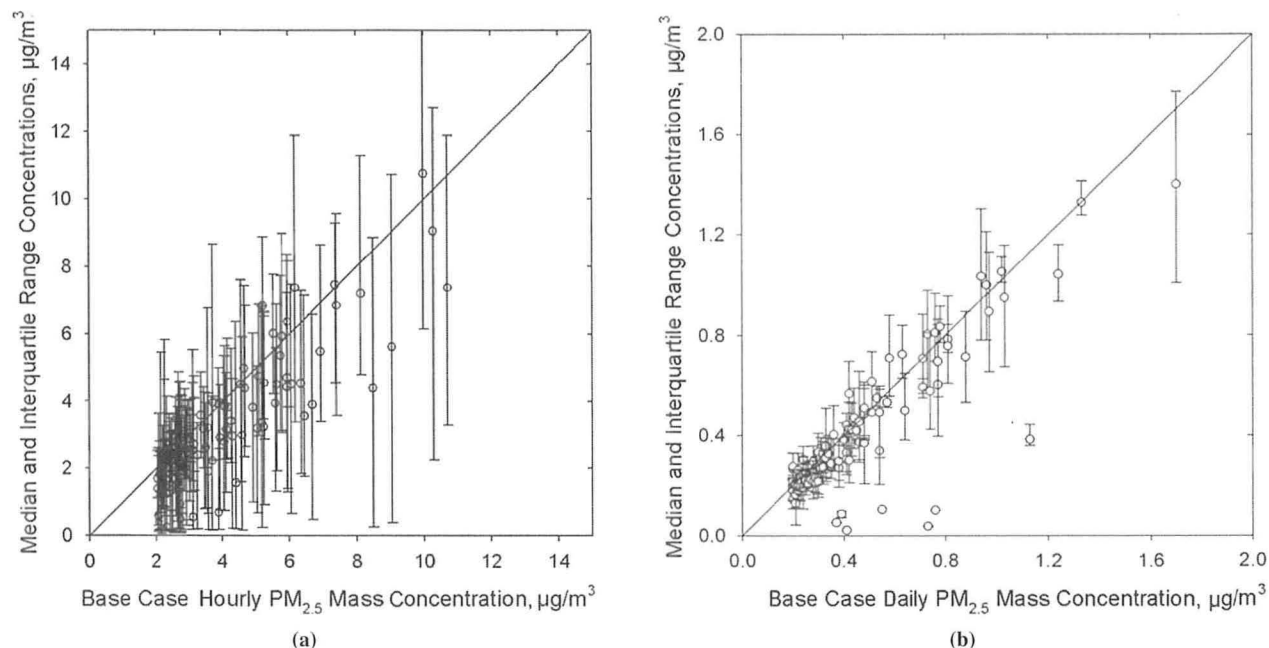


FIGURE 6 Modeled locomotive (a) hourly and (b) daily impacts at Dearborn station (2008) for base case (x-axis) and from Monte Carlo simulations that randomly placed the four switcher locomotives within their activity zones.

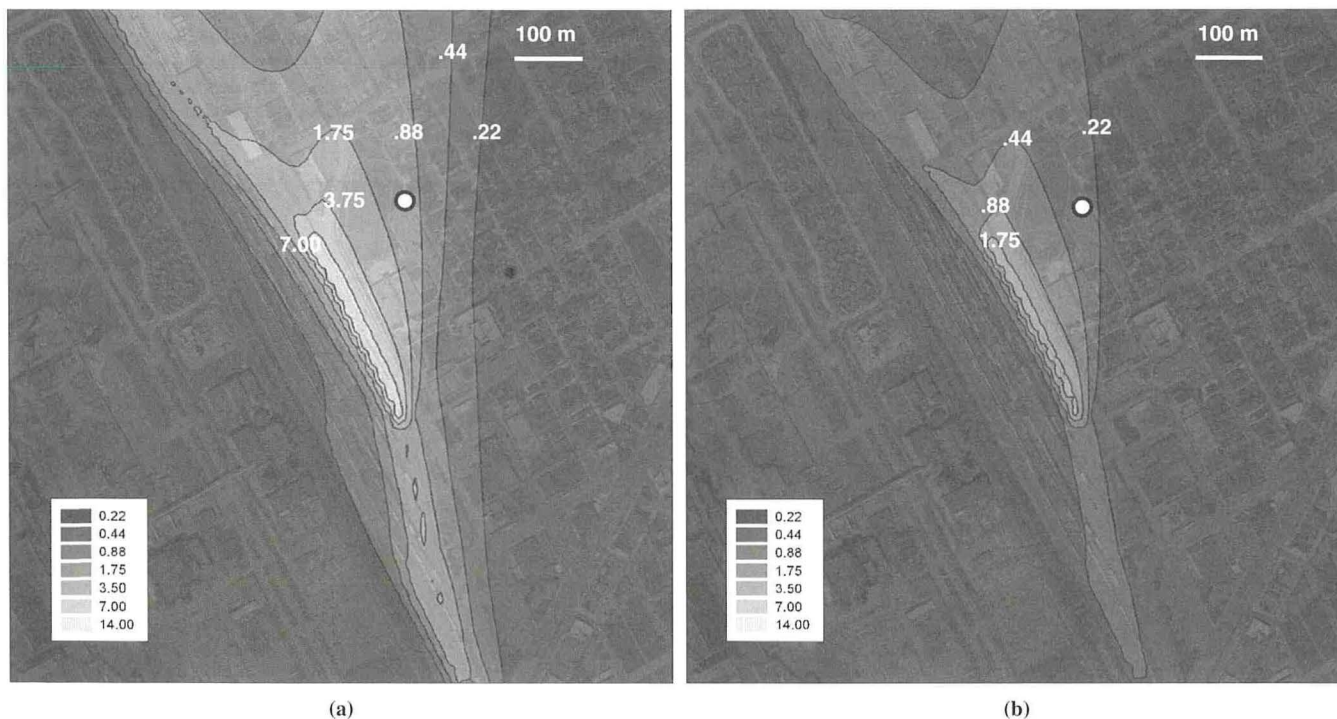


FIGURE 7 Modeled 24-h average  $\text{PM}_{2.5}$  mass impacts ( $\mu\text{g}/\text{m}^3$ ) using November 19, 2007, meteorology and emissions from (a) November 19, 2007, representing 95th percentile daily average concentration for rail yard impact at Dearborn station (white circle) for September to December 2007, and (b) November 17, 2008, representing emissions for same day of week, 1 year later, after all four switcher locomotives were replaced with GenSet technology.

at the Dearborn station from all locomotive activities were  $0.2 \mu\text{g}/\text{m}^3$  (annual average) and  $0.6 \mu\text{g}/\text{m}^3$  (90th percentile of daily averages) in 2008. However, the methodology used to model the Rougemere railyard provides a framework for modeling impacts at other railyards. Monte Carlo simulations for different switcher locomotive locations within the railyard demonstrate high variability in hourly average emissions impacts at the Dearborn monitor; the variability significantly decreases with increasing averaging time, and daily average impacts are well-represented by the spatial allocation used for the base case modeling.

## ACKNOWLEDGMENTS

The Midwest Rail Study was funded by a RARE Grant from the EPA Office of Research and Development to EPA Region 5. The authors gratefully acknowledge assistance from Loretta Lehrman, Monica Pagua, and other staff of EPA Region 5. The emission inventory development and dispersion modeling projects under the Midwest Rail Study were administered by the Lake Michigan Air Directors Consortium (LADCO). The authors are indebted to Michael Koerber for the detailed comments and valuable insights he has provided throughout this project. Donna Kenski and Mark Janssen of LADCO are also gratefully acknowledged. Jim Haywood of Michigan DEQ generously provided AERMOD files that he had prepared for an earlier local-scale analysis of sources affecting the Dearborn site. These files were helpful in guiding the dispersion modeling work. The cooperation of CSX to enable the emission inventory development is gratefully acknowledged.

## REFERENCES

1. Campbell, D. E., and E. M. Fujita. *Roseville Railyard Air Monitoring Project (RRAMP): Third Annual Report, Review and Summary of Year 3 (2007) Data*. Desert Research Institute, Reno, Nev., 2008.
2. Air Resources Board, California Environmental Protection Agency. *Railyard Health Risk Assessments and Mitigation Measures*. September 23, 2009. <http://www.arb.ca.gov/railyard/hra/hra.htm>. Accessed July 29, 2010.
3. Douglas, B. C., J. G. Heiken, and G. S. Rubenstein. Characterization and Breadth of Railyard-Specific Inventories. *Proc., USEPA 19th International Emission Inventory Conference*, U.S. Environmental Protection Agency, San Antonio, Tex., 2010. <http://www.epa.gov/ttn/chief/conference/ei19/index.html>. Accessed March 15, 2011.
4. Gould, G., and D. Niemeier. Review of Regional Locomotive Emission Modeling and the Constraints Posed by Activity Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2117, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 24–32.
5. Lindjhem, C. Intermodal Yard Activity and Emissions Evaluations. *Proc., USEPA 17th International Emission Inventory Conference*, U.S. Environmental Protection Agency, Portland, Ore., 2008.
6. Sangkapichai, M., J.-D. M. Saphores, S. G. Ritchie, S. You, and G. Lee. Analysis of PM and  $\text{NO}_x$  Train Emissions in the Alameda Corridor, California. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
7. Corfa, E., F. Maury, P. Segers, A. Fresneau, and A. Albergel. Short-Range Evaluation of Air Pollution near Bus and Railway Stations. *Science of the Total Environment*, Vol. 334–335, 2004, pp. 223–230.
8. Turner, J. R., V. Yadav, and S. N. Feinberg. *Data Analysis and Dispersion Modeling for the Midwest Rail Study (Phase I)*. Washington University, St. Louis, Mo., 2009.
9. Turner, J. R. *A Conceptual Model for Ambient Fine Particulate Matter over Southeast Michigan: High Concentration Days*. Version 1.0. Washington University, St. Louis, Mo. 2008.
10. *State Implementation Plan Submittal for Fine Particulate Matter ( $\text{PM}_{2.5}$ )*. Michigan Department of Environmental Quality, Lansing, 2008.
11. Heiken, J. G. *Rougemere Railyard Emission Inventory*. Sierra Research, Inc., Sacramento, Calif., 2009.
12. *Air Quality Technical Report for the ICTF Modernization Project, Long Beach, California*. Sierra Research, Inc., Sacramento, Calif., 2007.

Although the research described in this paper has been funded wholly or in part by the U.S. EPA, it has not been subjected to the EPA's required peer and policy review and therefore does not necessarily reflect the views of EPA, and no official endorsement should be inferred. CSX did not sponsor any portion of the Midwest Rail Study and did not formally review or endorse the project work plan or project deliverables.

The Passenger Rail Equipment and Systems Integration Committee peer-reviewed this paper.

# Integrated Optimization Model to Manage Risk of Transporting Hazardous Materials on Railroad Networks

Yung-Cheng Lai, Athaphon Kawprasert, Chen-Yu Lin, M. Rapik Saat, Chun-Hao Liang, and Christopher P. L. Barkan

Rail transport plays a key role in safely and economically moving hazardous materials from production to consumption points. As a result of heightened safety and security concerns, interest in all possible means of reducing the risk of transporting hazardous materials has intensified in recent years. Various approaches to railroad accident prevention—including infrastructure improvements, packaging enhancements, operational changes, and alteration of the route structure—are available. Operations research techniques have been applied to consider each approach individually, but no technique has integrated the approaches into a single model. This study introduced an integrated mathematical model to formally consider a combination of approaches to reduce risk. The framework enabled simultaneous consideration of route choices, tank car safety designs, and track maintenance to determine an optimal strategy that minimized risk and costs. Model formulation was provided in the form of nonlinear and mixed-integer programming. For illustration, a small-scale, hypothetical network flow of a hazardous material was considered. Numerical results showed that the optimal strategy could substantially reduce risk with a marginal increase in costs. The integrated model provided a framework for choosing the most effective risk-mitigation strategy for a particular rail network given various constraints. It could be applied to multiple types of commodities and adapted to address various questions for local, regional, or systemwide planning and decision making to provide the safest transportation possible given constrained resources. The framework would be particularly beneficial to rail carriers interested in how to best allocate safety and engineering resources to maximize safety.

Reducing railroad hazardous materials transportation risk has long been a priority for industry and government. As a result of security concerns and several fatal accidents this interest has intensified in recent years (1, 2). Particular attention has been directed toward over two dozen chemical products classified as toxic inhalation haz-

ard materials, including chlorine and ammonia. Efficient management of the risk posed by these materials requires an understanding of how different approaches may reduce risk and their relative cost-effectiveness both alone and in combination.

There are various approaches to hazardous materials transportation risk reduction. A variety of operations research techniques have been developed and applied to consider each approach individually, including consideration of hazardous materials transportation routing (3–6), improving transportation packaging (7–9), upgrading track infrastructure (10), and managing the operating speed of hazardous materials trains (11). Saat and Barkan developed a preliminary comparative analysis of the effect of tank car safety design versus alternative routing (12) and infrastructure improvements (13). However, the authors are unaware of research that considers and compares more than one approach to risk reduction simultaneously. Such comparison is important to objectively evaluate different approaches, possible interactive effects, and relative cost-effectiveness and to determine optimal strategies.

Each risk reduction strategy has characteristic benefit and cost functions. A release event is typically conditioned on a series of earlier events—a train accident or derailment, hazardous material car involvement, and hazardous material car damage and release. Each event has its own probability distribution, which in turn affects the result of the risk equation. Lowering any of the terms in this equation will reduce risk, but the form and extent of the reduction associated with each term varies. Different risk reduction strategies affect the terms differently. For example, packaging enhancement involving tank car design improvement reduces the conditional probability of release from a tank car involved in an accident but does not reduce accident rates. Conversely, upgrading track infrastructure offers reduction in accident rates but does not affect the conditional probability of release from a tank car involved in an accident provided that operating speed remains the same. However, there are interactive effects among the terms that affect the cost-benefit analysis, which complicates comparison among different risk reduction strategies. The multiplicative form of the risk equation means that the benefit associated with a particular risk reduction strategy is affected by changes to others terms in the equation. Thus improving packaging reduces the benefit derived from improving infrastructure, and vice versa, but the cost associated with each of these strategies is unchanged. Consequently, implementing a risk reduction strategy may reduce the cost-effectiveness of other strategies.

This study presents an integrated risk management framework model to provide a means of choosing the most effective set of risk mitigation strategies for a particular rail network. The model is first formulated using nonlinear programming (NLP) and converted into

Y.-C. Lai, C.-Y. Lin, and C.-H. Liang, Railway Technology Research Center, Department of Civil Engineering, National Taiwan University, Civil Engineering Building, Room 313, No. 1, Roosevelt Road, Section 4, Taipei 10617, Taiwan. A. Kawprasert, M. R. Saat, and C. P. L. Barkan, Rail Transportation and Engineering Center, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 North Mathews Avenue, Urbana, IL 61801. Current affiliation for A. Kawprasert: Civil Engineering Department, State Railway of Thailand, 1 Rongmuang Road, Pathumwan, Bangkok 10330, Thailand. Corresponding author: Y.-C. Lai, yclai@ntu.edu.tw.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 115–123.  
DOI: 10.3141/2261-13

a mixed-integer programming (MIP) problem. For illustration, a small-scale, hypothetical network flow of a hazardous material is considered. The model is used to determine the optimal combination of number of shipments, tank car utilization, and track classes to be maintained on each link that minimizes total cost, given capacity, and demand constraints. The flexibility of the model framework is illustrated by considering different investment scenarios: upgrading track infrastructure, using a more robust tank car, and a combination of both.

## ELEMENTS OF RISK ANALYSIS AND OPTIMIZATION FRAMEWORK

Earlier work by Lai et al. identified optimal train routing over a network with capacity constraints and traffic heterogeneity to minimize transportation and track maintenance costs (6). Lai et al. also described a model that considers the trade-off between transportation and track maintenance costs and determines the optimal assignment for track class given traffic demand and track maintenance budget (14). These models, however, do not incorporate the safety aspect, in particular, the risk involved with transportation of hazardous materials.

In the present study, operations research techniques using network flow mathematical modeling are combined with a quantitative risk analysis model to develop an integrated framework to consider various risk management strategies. The model determines an optimal set of routes, track classes to be maintained, and tank car types that optimize both financial and safety impacts of hazardous materials transportation at the operational level.

Figure 1 depicts the conceptual diagram for the model for the transportation network assignment of hazardous materials. Input parameters include track infrastructure characteristics, tank car safety design features, and product characteristics. Two additional parameters, traffic demand and track capacity, represent the constraints of the optimization problem. The optimization framework is designed to deliver three types of output depending on the objective

of interest: hazardous material traffic flows, tank car type, and track classes to be maintained.

This initial stage of model development focuses on the operating cost components related to routes, track classes, and car types. Capital cost components are excluded, but they can easily be added if data are available. The problem is simplified by omitting some risk parameters, including population distribution along rail lines; chemical-specific hazard exposure, which varies with various factors including toxicity (15, 16); population densities; and train operating speed, which affects the conditional probability of release (17–20). The goal of this paper is to illustrate the fundamental methodology of the integrated model. Additional parameters not considered here will be addressed in future work as the model is further enhanced.

## MATHEMATICAL MODEL FORMULATION

### Nonlinear Programming Model

The following notation is used in the NLP model:

- $i$  and  $j$  = indices representing nodes;
- $N$  = set of all nodes;
- $A$  = set of all existing arcs  $i, j$ ;
- $K$  = set of  $k$ , where  $k$  corresponds to the  $k$ th origin–destination (O-D) pairs of node  $(s_1, e_1), (s_2, e_2), \dots, (s_k, e_k)$ , in which  $s_k$  and  $e_k$  denote the origin and destination, respectively, of the  $k$ th O-D pair;
- $Q$  = set of  $q$ , where  $q$  represents track class;
- $\delta^+(j)$  = station  $j$  serving as departure station;
- $\delta^-(j)$  = station  $j$  serving as arrival station;
- $A_k$  = additional number of cars for O-D pair  $k$  if the enhanced-safety tank car is selected;
- $C_{ij}$  = transportation cost per car mile on arc  $i, j$ ;
- $D_k$  = demand of O-D pair  $k$  (in number of cars);
- $H_{ij}^q$  = maintenance cost on arc  $i, j$  with track class  $q$ ;

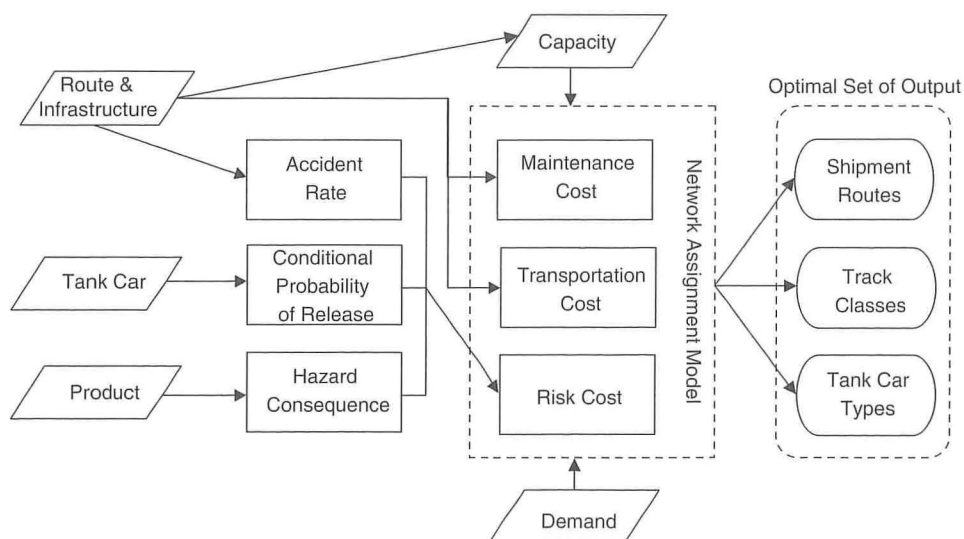


FIGURE 1 Conceptual diagram showing input–output of model for transportation network assignment of hazardous materials.

- $U_{ij}^q$  = capacity (in number of cars) on arc  $i, j$  with track class  $q$ ;  
 $R_{ij}^q$  = release risk per carload (in monetary value) on arc  $i, j$  with track class  $q$ ; and  
 $W_{ij}^q$  = reduced risk per carload (in monetary value) on arc  $i, j$  with track class  $q$  if enhanced tank car is selected.

Enhanced-safety tank cars are usually heavier than the baseline tank car, resulting in more cars needed for the same quantity transported. Although a higher track class also requires higher maintenance cost, it offers better safety because it is associated with a lower accident rate (21), and it increases the fluidity of the section with higher maximum speed (22), resulting in better capacity.

The NLP model has three sets of decision variables. The first variable ( $x_{ij}^k$ ) is a positive integer representing the number of cars running on arc  $i, j$  corresponding to O-D pair  $k$ . The second variable ( $y_{ij}^q$ ) is a binary variable that determines whether track class  $q$  is assigned on arc  $i, j$ . The third variable ( $z_k$ ) is also a binary variable to determine the tank-car enhancement for the  $k$ th O-D pair. The network assignment model is expressed in NLP form as follows:

$$\begin{aligned} \min \quad & \sum_{(i,j) \in A} \sum_{q \in Q} \sum_{k \in K} H_{ij}^q y_{ij}^q x_{ij}^k + \sum_{(i,j) \in A} \sum_{k \in K} C_{ij} x_{ij}^k \\ & + \sum_{(i,j) \in A} \sum_{q \in Q} \sum_{k \in K} (R_{ij}^q - W_{ij}^q z_k) y_{ij}^q x_{ij}^k \end{aligned} \quad (1)$$

subject to

$$\sum_{k \in K} (x_{ij}^k + x_{ji}^k) \leq \sum_{q \in Q} U_{ij}^q y_{ij}^q \quad \forall (i, j) \in A, (i < j) \quad (2)$$

$$\sum_{q \in Q} y_{ij}^q = 1 \quad \forall (i, j) \in A, (i < j) \quad (3)$$

$$\sum_{j \in \delta^-(j)} x_{ij}^k - \sum_{j \in \delta^+(j)} x_{ji}^k = \begin{cases} D_k + A_k z_k & \text{if } i \in s_k \\ -D_k - A_k z_k & \text{if } i \in e_k \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in N, k \in K \quad (4)$$

and

$$\begin{aligned} x_{ij}^k &\in \text{positive integer} & \forall (i, j) \in A, k \in K \\ y_{ij}^q &\in \{0, 1\} & \forall (i, j) \in A, k \in K \\ z_k &\in \{0, 1\} & \forall k \in K \end{aligned} \quad (5)$$

The objective function in Equation 1 minimizes the sum of total track maintenance cost ( $\sum_{(i,j) \in A} \sum_{q \in Q} \sum_{k \in K} H_{ij}^q y_{ij}^q x_{ij}^k$ ), total transportation cost ( $\sum_{(i,j) \in A} \sum_{k \in K} C_{ij} x_{ij}^k$ ), and total risk cost [ $\sum_{(i,j) \in A} \sum_{q \in Q} \sum_{k \in K} (R_{ij}^q - W_{ij}^q z_k) y_{ij}^q x_{ij}^k$ ]. Equation 2 is the capacity constraint. Equation 3 ensures that only one track class is assigned for each arc  $i, j$ . Finally, Equation 4 is the flow conservation constraint. That is, if the enhanced tank car is selected, then additional numbers of tank cars ( $A_k$ ) will be added. This model determines the optimal assignments of the number of shipments, type of tank cars used, and track classes to be maintained while minimizing the total cost comprising track maintenance cost, transportation cost, and risk cost.

## Mixed-Integer Programming Model

To ensure the global optimal solution, the nonlinear model is converted into a linear form by using MIP. The following notation is used in the linear model:

- $i$  = index referring to the starting node of an arc;  
 $j$  = ending node of an arc;  
 $k$  =  $k$ th O-D pairs of node  $(s_1, e_1), (s_2, e_2), \dots, (s_k, e_k)$ , in which  $s_k$  and  $e_k$  denote the origin and destination, respectively, of the  $k$ th O-D pair;  
 $q$  = track class;  
 $t$  = type of tank car (baseline or enhanced);  
 $T$  = set of the two car types;  
 $D_{kt}$  = demand expressed as number of shipments for O-D pair  $k$  and type  $t$  tank car;  
 $V$  = set of  $v$ , where  $v$  is an index representing traffic composition where each  $v$  refers to a specific combination of car types [e.g.,  $v = (N_1, N_2) = (3, 6)$  means there are three baseline tank cars and six enhanced tank cars];  
 $N_t^v$  = number of type  $t$  tank cars in traffic composition  $v$ ;  
 $C_{ij}$  = transportation cost per carload on arc  $i, j$ ;  
 $H_{ij}^{vq}$  = maintenance cost of arc  $i, j$  with track class  $q$  and traffic composition  $v$ ;  
 $R_{ij}^{vq}$  = unit cost of release risk on arc  $i, j$  with track class  $q$  and traffic composition  $v$ ; and  
 $U_{ij}^q$  = capacity (in number of cars) on arc  $i, j$  with track class  $q$ .

This model has three sets of decision variables. The first variable ( $x_{ij}^{kt}$ ) indicates the number of cars running on arc  $i, j$  for O-D pair  $k$  and car type  $t$ . The second variable ( $y_{ij}^{vq}$ ) is a binary variable used to determine the traffic composition ( $v$ ) of arc  $i, j$  under particular track class  $q$ . The third variable ( $z_{kt}$ ) is also a binary variable to determine the tank car type  $t$  for O-D pair  $k$ .

The linear optimization model is formulated as follows:

$$\min \quad \sum_{(i,j) \in A} \sum_{v \in V} \sum_{q \in Q} H_{ij}^{vq} y_{ij}^{vq} + \sum_{(i,j) \in A} \sum_{k \in K} \sum_{t \in T} C_{ij} x_{ij}^{kt} + \sum_{(i,j) \in A} \sum_{v \in V} \sum_{q \in Q} R_{ij}^{vq} y_{ij}^{vq} \quad (6)$$

subject to

$$\sum_{k \in K} \sum_{t \in T} (x_{ij}^{kt} + x_{ji}^{kt}) \leq \sum_{v \in V} \sum_{q \in Q} U_{ij}^q y_{ij}^{vq} \quad \forall (i, j) \in A, (i < j) \quad (7)$$

$$\sum_{v \in V} \sum_{q \in Q} y_{ij}^{vq} = 1 \quad \forall (i, j) \in A, (i < j) \quad (8)$$

$$\sum_{k \in K} (x_{ij}^{kt} + x_{ji}^{kt}) \leq \sum_{v \in V} \sum_{q \in Q} N_t^v y_{ij}^{vq} \quad \forall (i, j) \in A, (i < j), t \in T \quad (9)$$

$$\sum_{j \in \delta^-(j)} x_{ij}^{kt} - \sum_{j \in \delta^+(j)} x_{ji}^{kt} = \begin{cases} D_{kt} z_{kt} & \text{if } i \in s_k \\ -D_{kt} z_{kt} & \text{if } i \in e_k \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in N, k \in K, t \in T \quad (10)$$

$$\sum_{t \in T} z_{kt} = 1 \quad \forall k \in K \quad (11)$$

and

$$\begin{aligned} x_{ij}^{kt} &\in \text{positive integer} & \forall (i, j) \in A, k \in K, t \in T \\ y_{ij}^{vq} &\in \{0, 1\} & \forall (i, j) \in A, v \in V, q \in Q \\ z_{kt} &\in \{0, 1\} & \forall k \in K, t \in T \end{aligned} \quad (12)$$

The objective function in Equation 6 minimizes the sum of total track maintenance cost ( $\sum_{(i,j) \in A} \sum_{v \in V} \sum_{q \in Q} H_{ij}^{vq} y_{ij}^{vq}$ ), total transportation cost ( $\sum_{(i,j) \in A} \sum_{k \in K} \sum_{t \in T} C_{ij} x_{ij}^{kt}$ ), and total risk cost ( $\sum_{(i,j) \in A} \sum_{v \in V} \sum_{q \in Q} R_{ij}^{vq} y_{ij}^{vq}$ ). Equation 7 is the capacity constraint. Equation 8 ensures that only one traffic composition and track class are selected for arc  $i, j$ . Equation 9 is the linking constraint between  $x_{ij}^{kt}$  and  $y_{ij}^{vq}$  to maintain the consistency of car types assigned. Equation 10 is the flow conservation constraint. Finally, Equation 11 ensures that only one type of tank car is assigned to a particular O-D pair  $k$ .

### Additional Constraints on Traffic Flow

In some instances, it may be preferable to assign traffic from the same O-D pair to the same route. To implement this routing strategy, a new binary decision variable ( $w_{ij}^{kt}$ ) and the following two constraints should be added to the original formulation:

$$\sum_{j \in \delta^+(i)} \sum_{t \in T} w_{ij}^{kt} = 1 \quad \forall i \in N, k \in K \quad (13)$$

$$x_{ij}^{kt} \leq M w_{ij}^{kt} \quad \forall (i, j) \in A, k \in K, t \in T \quad (14)$$

The binary variable ( $w_{ij}^{kt}$ ) equals 1 if arc  $i, j$  is selected for O-D pair  $k$  and car type  $t$ .  $M$  is a large number to ensure that  $x_{ij}^{kt}$  can have a valid value. Equations 13 and 14 ensure that shipments from the same O-D pair will be combined together on the same route during the traffic assignment.

### CASE STUDY

This case study considers a hypothetical transportation network comprising nine nodes, 12 links, and four O-D pairs (Figure 2 and Table 1). The following table shows the input parameters for traffic data:

O-D Pair	Daily Shipment Volume (gal)
AI	2,100,000
GC	3,600,000
DI	3,900,000
BI	5,700,000

Two cases (with and without additional traffic flow constraints) are defined and optimized. Route lengths are shown on each link, and FRA track classes are indicated by numbers in *italics*. Among these nodes, E is considered as a city linked by Class 4 track, C as a medium-sized town linked by Class 3 track with moderate route length, and G as a small village with Class 3 track and longer route length. Track capacity is assumed, and O-D flows are considered of a particular hazardous material using two tank car types (Table 1).

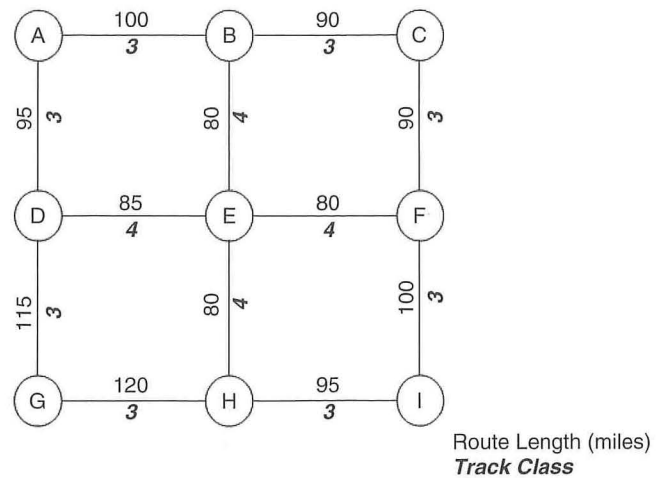


FIGURE 2 Hypothetical network.

For illustration, only one type of product is considered, but the model can be adapted to accommodate multiple commodities.

To estimate risk, the following equation is used:

$$R = P_1 \times P_2 \times M \times L \times C \quad (15)$$

where

$R$  = risk of hazardous material release (millions of dollars),  
 $P_1$  = track class-specific accident rate (cars derailed per car mile),  
 $P_2$  = conditional probability of release given that a tank car is derailed in an accident,  
 $M$  = shipments (carloads),  
 $L$  = mileage, and  
 $C$  = average consequence cost per release incident (millions of dollars).

The model uses the derailment rates per car mile developed by Anderson and Barkan (21) and the conditional probability of release given that a tank car is derailed in an accident as developed by Treichel et al. (19) and shown in the following table:

Tank Car Type	Capacity (gal)	Conditional Probability of Release
Baseline	30,000	.3527
Enhanced	28,947	.2681

TABLE 1 Input Parameters for Case Study: Track Class Information

Track Class	Accident Rate (cars derailed per car mile)	Track Capacity (cars per day)	Coefficients for Track Maintenance Cost Function (23)	
			$\alpha$	$\beta$
3	$300 \times 10^{-9}$	1,200	651.6	51.5
4	$77 \times 10^{-9}$	1,700	811.7	57.9
5	$42 \times 10^{-9}$	2,300	935.9	62.8

The model does not consider the effects of train length and schedule. The results may change with more specific train length, train capacity, and scheduling constraints.

The consequence analysis is simplified by neglecting population distribution along the rail lines, and an average consequence cost of one million dollars per release incident on all track classes is assumed. Unit transportation cost is determined using data from the Association of American Railroads (24), resulting in \$0.55 per car mile. For the maintenance cost corresponding to a particular track class, the formulation developed in Lai et al. (14) using data from Zarembski et al. (25) is adopted. Capital investment cost is excluded in this case study but can be added if data are available. The maintenance cost function is as follows:

$$MC = \alpha X + \beta \quad (16)$$

where

MC = average maintenance cost (millions of dollars per mile),  
 X = tonnage (million gross tons), and  
 $\alpha$  and  $\beta$  = model coefficients.

The railroad network is assumed to use wood ties and has predominantly tangent or moderate curvature track alignment as defined by Zarembski et al. (25). The values of  $\alpha$  and  $\beta$  are given in Table 1.

To illustrate the potential application of the hazardous materials transportation network assignment model, four scenarios are considered for each case as follows:

- Baseline (neither tank cars nor track infrastructure are upgraded),
- Infrastructure upgrade only,
- Tank car upgrade only, and
- Combined upgrades (both track infrastructure and tank car upgrades are allowed).

### Case I. Without Constraint on Traffic Flow

The model is used to determine the flows of hazardous material under each scenario with the objective of minimizing total cost. The model is formulated using the general algebraic modeling system and solved using CPLEX (26). In Figures 3 and 4 the first and second numbers in parentheses represent daily shipments (carloads) made

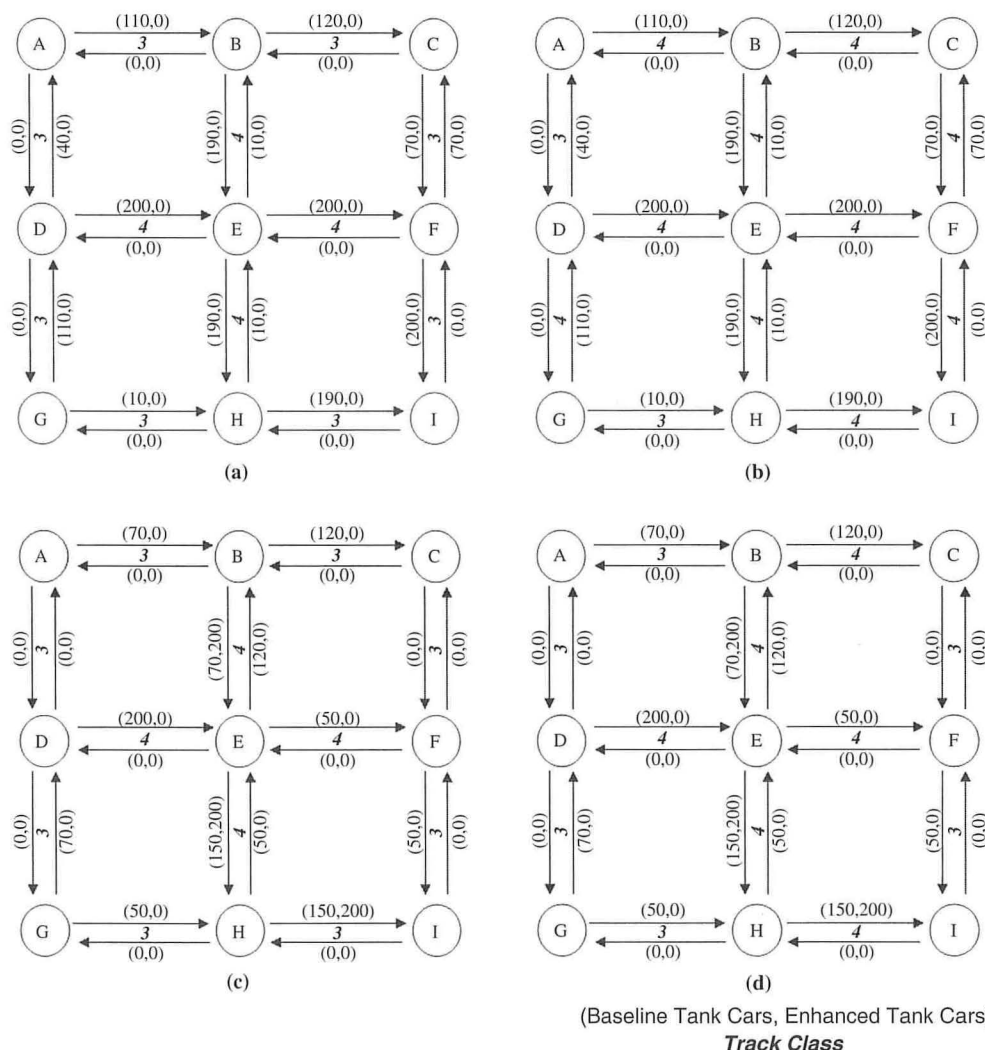


FIGURE 3 Optimal routing and track class assignment in Case I for (a) baseline, (b) infrastructure upgrade only, (c) tank car upgrade only, and (d) combined upgrades.

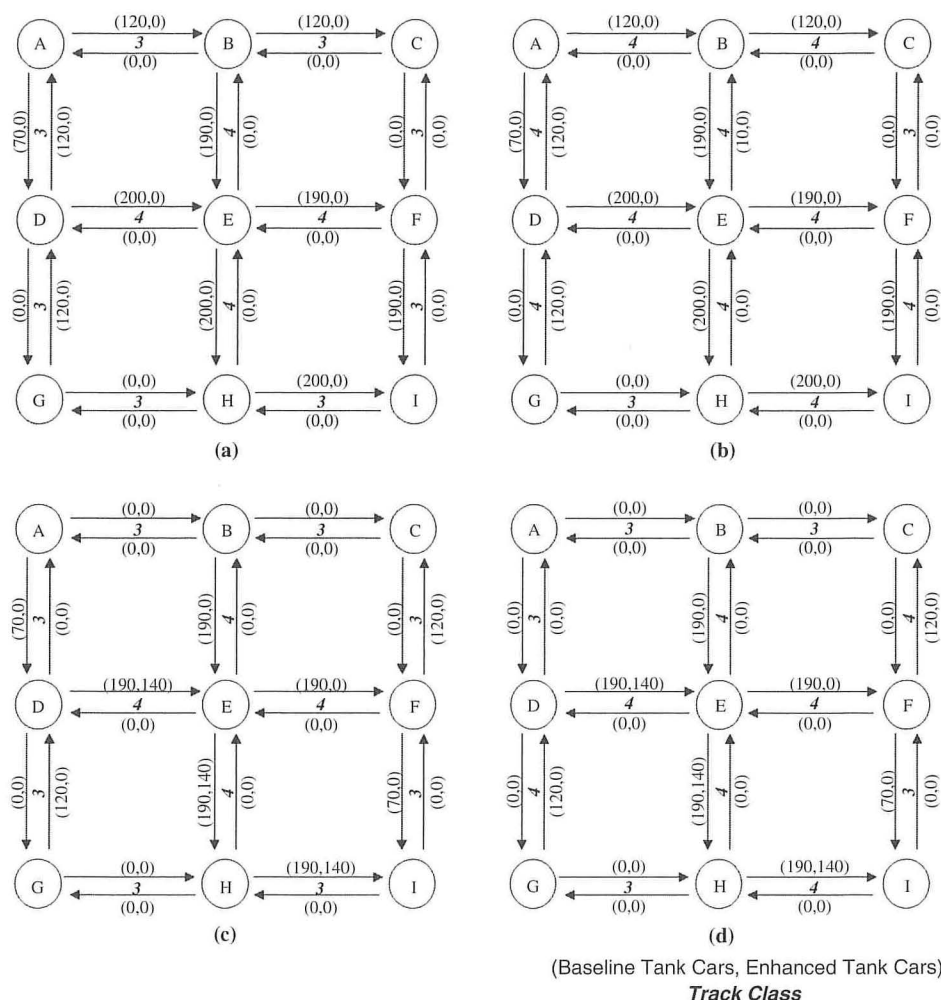


FIGURE 4 Optimal routing and track class assignment in Case II for (a) baseline, (b) infrastructure upgrade only, (c) tank car upgrade only, and (d) combined upgrades.

using baseline and enhanced tank cars, respectively. The italicized numbers represent the track class to be maintained.

The solution for the baseline scenario (Figure 3a) represents the flows of hazardous material on existing infrastructure using the baseline tank car. In the second scenario, in which only track infrastructure upgrade is allowed (Figure 3b), the model suggested upgrading six of eight Class 3 tracks to Class 4, indicating that the maintenance cost will be noticeably higher than the baseline scenario. In the third scenario, in which only tank car upgrade is allowed (Figure 3c), traffic flows on some links are eliminated. The fourth scenario represents the case in which both track and rolling stock can be upgraded (Figure 3d). The total cost for the latter scenario is expected to be the lowest of the four because of the greater flexibility in choosing upgrade options while reducing risk. Although the enhanced tank car has a lower conditional probability of release, more shipments are needed because of lower capacity compared with the baseline tank car.

Table 2 shows that the total cost of the baseline scenario is the highest and the combined upgrades scenario, the lowest. Upgrading track infrastructure requires a higher cost for track maintenance (\$1.98 million or 7.38% greater than baseline), but more than half (59.24%) of the risk can be reduced. Upgrading

tank cars offers the smallest reduction in risk, but the reduction in transportation cost is the greatest. A combination of infrastructure and tank car upgrades is the optimal scenario associated with the greatest reduction in total costs. For the problem considered, 44.93% of risk can be reduced with a 2.37% increase in maintenance cost.

### Case II. With Constraint on Traffic Flow

In some instances, assigning traffic between the same O-D pair to the same route may be preferable. For this particular case, special routing requirements are implemented for the same problem (Figure 4 and Table 3).

Comparison of the costs among different scenarios shows almost the same trend as the unconstrained traffic flow case (Case I). Most of the costs become slightly higher compared with Case I because of the additional routing constraints. As expected, the traffic, track class assignment, and tank car selection results differ from those in Case I. This illustrates the dynamic nature of the problem and the potential insights that can be gained through application of this integrated optimization framework.

TABLE 2 Comparison of Annual Cost Components Without Constraint on Traffic Flow

Scenario	Maintenance		Transportation		Risk		Total	
	Cost (\$ million)	Reduction (%)	Cost (\$ million)	Reduction (%)	Cost (\$ million)	Reduction (%)	Cost (\$ million)	Reduction (%)
Baseline	26.84	—	31.14	—	4.12	—	62.10	—
Infrastructure upgrade	28.82	(7.38)	31.14	0.00	1.68	59.24	61.64	0.75
Tank car upgrade	26.95	(0.42)	30.98	0.51	3.29	20.22	61.22	1.42
Combined upgrades	27.48	(2.37)	30.98	0.51	2.27	44.93	60.72	2.22

NOTE: — = not applicable.

### Case III. Minimizing Risk of Hazardous Material Transportation

The optimization framework described earlier minimizes maintenance cost, transportation cost, and risk cost simultaneously. In this case, risk cost is considered as the only component in the objective function. The optimal results suggest the use of enhanced tank cars and upgrading of all segments with hazardous materials traffic to Class 5 (Table 4). Compared with the full model, which incorporates maintenance cost, transportation cost, and risk cost, the resulting risk cost is much lower, while the total cost is much higher than the previous two cases with consideration of the overall cost (including maintenance cost, transportation cost, and risk cost). If risk cost is excluded, the results will be very similar to those in Cases I and II because risk cost shares only a small portion of the overall cost (Tables 2 and 3).

### DISCUSSION OF MODEL RESULTS

Railroad hazardous materials transportation safety depends on the design and condition of the railroad infrastructure and operating practices on the routes they travel (27) and the damage resistance of the tank cars transporting them. In addition to routing, improvements to either infrastructure or rolling stock, or both, have some potential to enhance safety, but there are different functional relationships between the cost and safety benefit for each. In different situations, investing in either or both may be the most efficient means of improving safety. This study addresses these elements individually and simultaneously.

The mathematical framework presented in this paper allows better consideration of a combination of different risk reduction strate-

gies that potentially offer the greatest safety benefit at the lowest total cost. Besides incorporating the route-specific consequence elements, the model can be implemented to address a real-world rail network with complete track segment characterization (with variables affecting risk and maintenance cost) and O-D-level traffic information for commodities.

While improving infrastructure is generally more costly than other risk reduction strategies, it also reduces the risk of accidents involving all types of hazardous materials as well as other products traveling over the affected section. However, the benefit is isolated to those locations where the infrastructure was upgraded. In contrast, improving tank car safety design only affects risk for the products being transported, but that benefit is realized everywhere the improved tank cars travel in the network. Meanwhile, routing decisions often involve a complex set of other interacting factors that both increase and reduce safety and risk. Consequently, the net effect will be highly route and commodity specific and will depend on the particular combination of circumstances involved.

The hypothetical case study considered here demonstrates a potential reduction in transportation risk (44% lower than baseline) under combined, optimized strategies of routing, tank car safety design enhancement, and track maintenance with a slight increase in track maintenance cost (2.4% increase from baseline) (Table 4). Both the tank-car-upgrade-only and infrastructure-upgrade-only scenarios improve the performance from the baseline scenario, but integration of both options provides the best solution. Different network conditions or constraints would result in a different optimal result, so an integrated optimization framework such as the one described in this paper is necessary to improve hazardous materials transportation safety in the most efficient manner possible.

The case study presented here is an example of transportation of a single product. It does not consider the effect of train speed on

TABLE 3 Comparison of Annual Cost Components with Constraint on Traffic Flow

Scenario	Maintenance		Transportation		Risk		Total	
	Cost (\$ million)	Reduction (%)	Cost (\$ million)	Reduction (%)	Cost (\$ million)	Reduction (%)	Cost (\$ million)	Reduction (%)
Baseline	26.86	—	31.33	—	4.21	—	62.39	—
Infrastructure upgrade	28.90	(7.61)	31.33	0.00	1.55	63.23	61.77	0.99
Tank car upgrade	26.96	(0.39)	31.01	1.02	3.35	20.35	61.32	1.72
Combined upgrades	28.05	(4.46)	31.01	1.02	1.84	56.36	60.90	2.39

TABLE 4 Comparisons of Annual Cost Components of Minimal Risk Cases Without and With Constraint on Traffic Flow

Scenario	Maintenance		Transportation		Risk		Total	
	Cost (\$ million)	Reduction (%)	Cost (\$ million)	Reduction (%)	Cost (\$ million)	Reduction (%)	Cost (\$ million)	Reduction (%)
Without Constraint on Traffic Flow								
Baseline	26.84	—	31.14	—	4.12	—	62.10	—
Infrastructure upgrade	32.22	(20.05)	31.14	0.00	0.84	79.66	64.20	(3.38)
Tank car upgrade	27.18	(1.28)	32.50	(4.38)	3.03	26.48	62.72	(0.99)
Combined upgrades	32.54	(21.23)	32.69	(5.00)	0.70	82.98	65.93	(6.17)
With Constraint on Traffic Flow								
Baseline	26.86	—	31.33	—	4.21	—	62.39	—
Infrastructure upgrade	31.39	(16.89)	31.33	0.00	0.84	79.94	63.56	(1.88)
Tank car upgrade	27.06	(0.78)	31.67	(1.09)	3.06	27.24	61.79	0.96
Combined upgrades	31.75	(18.23)	33.11	(5.70)	0.71	83.12	65.58	(5.11)

conditional probability of release (17–20), nor does it take into account population distribution along the route. The model did account for transportation cost and track maintenance and renewal cost. A constant consequence cost and the same unit transportation costs were assumed for all the links in the hypothetical network. The model does not consider the time value of money and assets, such as depreciation and amortization of rolling stock or increases in infrastructure improvement cost caused by the interest rate. If these data are available, more accurate costs that vary with train speed or track class and the change over time could be incorporated. Future research may also consider the differential costs of different tank car types in addition to one-time investment costs in infrastructure and rolling stock. Speed reduction could also be evaluated as a risk management strategy. The model framework developed here can be modified to accommodate all of these additional factors. For example, the conditional probability of release that is dependent on train speed and track segment-specific characteristics and exposure to population could be used in the risk model. Multiple products can be modeled by enlarging the index representing traffic composition ( $v$ ). Another index representing time can be added to take into account time value of money and assets.

In this study, the optimal combination of different risk reduction strategies was identified on the basis of the assumption of a single decision maker. It was also assumed that the associated costs and benefits are incurred and gained by the same decision maker. In practice, railroad hazardous materials transportation involves a number of different entities including railroads, shippers, consignees, and car owners. Different parties are subject to different liabilities, although railroads generally assume principal liability in accidents unless it can be shown that the accident or release was the fault of one of these other parties. The additional costs for enhanced-safety tank cars are generally incurred by the car owners or shippers (or both), but the benefit of the reduction in risk is generally accrued by the railroad. The optimization model in this paper provides a globally optimal solution if all entities behave in a systematically rational manner with the same risk minimization effectiveness goal. However, with one set of parties paying for the enhancements and another set receiving the benefits, the potential exists for conflicting objectives and constraints. These add to the complexity of optimizing decision making and should be considered when using the model to consider different risk management strategies.

## CONCLUSION

An integrated risk management framework with a network assignment model is presented that determines an optimal combination of different strategies to minimize the costs of transportation, track maintenance, and risk in hazardous materials transportation. The model advances understanding of how to most efficiently and effectively manage risk and provides guidance for tactical and strategic operational control, infrastructure and vehicle design, and maintenance for public and private sector policy making.

## ACKNOWLEDGMENTS

This work was part of a collaborative research project of the National Taiwan University and the University of Illinois at Urbana–Champaign on railroad transportation safety. The research that made this work possible was conducted by students who were supported by grants or gifts to the University of Illinois at Urbana–Champaign from the Association of American Railroads, BNSF Railway, CN Research Fellowships in Railroad Engineering, Dow Chemical, Monsanto, Railway Supply Institute, and others. The authors are sincerely grateful to these organizations for their support.

## REFERENCES

1. U.S. Department of Transportation. *Hazardous Materials: Enhancing Rail Transportation Safety and Security for Hazardous Materials Shipments; Railroad Safety Enforcement Procedures; Enforcement, Appeal and Hearing Procedures for Rail Routing Decisions; Final Rules*. 49 CFR Parts 172, 174, and 209. *Federal Register*, Vol. 73, No. 229, 2008.
2. U.S. Department of Transportation. *Positive Train Control Systems*. 49 CFR Parts 229, 234, 235, and 236. *Federal Register*, Vol. 75, No. 10, 2010.
3. Glickman, T. S. Rerouting Railroad Shipments of Hazardous Materials to Avoid Populated Areas. *Accident Analysis and Prevention*, Vol. 15, No. 5, 1983, pp. 329–335.
4. Glickman, T. S., E. Erkut, and M. S. Zschocke. The Cost and Risk Impacts of Rerouting Railroad Shipments of Hazardous Materials. *Accident Analysis and Prevention*, Vol. 39, 2007, pp. 1015–1025.
5. Kawprasert, A., and C. P. L. Barkan. Effects of Route Rationalization on Hazardous Materials Transportation Risk. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2043, Trans-

- portation Research Board of the National Academies, Washington, D.C., 2008, pp. 65–72.
6. Lai, Y.-C., M. H. Dangler, C.-E. Hsu, and P.-C. Chiang. Optimizing Train Network Routing with Heterogeneous Traffic. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2159*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 69–76.
  7. Barkan, C. P. L., T. S. Glickman, and A. E. Harvey. Benefit–Cost Evaluation of Using Different Specification Tank Cars to Reduce the Risk of Transporting Environmentally Sensitive Chemicals. In *Transportation Research Record 1313*, TRB, National Research Council, Washington, D.C., 1991, pp. 33–43.
  8. Barkan, C. P. L. Improving the Design of Higher-Capacity Railway Tank Cars for Hazardous Materials Transport: Optimizing the Trade-Off Between Weight and Safety. *Journal of Hazardous Materials*, Vol. 160, 2008, pp. 122–134.
  9. Saat, M. R. *Optimizing Railroad Tank Car Safety Design to Reduce Hazardous Materials Transportation Risk*. PhD dissertation. University of Illinois at Urbana–Champaign, 2009.
  10. Kawprasert, A., and C. P. L. Barkan. Options for Rail Route Infrastructure Improvement to Reduce Hazardous Materials Transportation Risk. Presented at ASME-IEEE-ASCE-AREMA-TRB Joint Rail Conference, University of Illinois at Urbana–Champaign, 2010.
  11. Kawprasert, A. *Quantitative Analysis of Options to Reduce Risk of Hazardous Materials Transportation by Railroad*. PhD dissertation. University of Illinois at Urbana–Champaign, 2010.
  12. Saat, M. R., and C. P. L. Barkan. The Effect of Rerouting and Tank Car Safety Design on the Risk of Rail Transport of Hazardous Materials. *Proc., 7th World Congress on Railway Research*, Montreal, Quebec, Canada, 2006.
  13. Saat, M. R., and C. P. L. Barkan. Tank Car Safety Design Versus Infrastructure Improvements in Reducing Hazardous Materials Transportation Risks. Presented at INFORMS Annual Meeting, Pittsburgh, Pa., 2006.
  14. Lai, Y.-C., M. H. Dangler, and S.-Y. Lin. Optimization of the Railroad Network Track Class. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.
  15. *Emergency Response Guidebook*. Pipeline and Hazardous Materials Safety Administration, U.S. Department of Transportation, 2008.
  16. Brown, D. F., H. M. Hartmann, W. A. Freeman, and W. D. Haney. *Development of the Table of Initial Isolation and Protective Action Distances for the 2008 Emergency Response Guidebook*. Publication ANL/DIS-09-2. Argonne National Laboratory, U.S. Department of Energy, 2009.
  17. Nayak, P. R., D. B. Rosenfield, and J. H. Hagopian. *Event Probabilities and Impact Zones for Hazardous Materials Accidents on Railroads*. Report DOT/FRA/ORD-83/20. FRA, U.S. Department of Transportation, 1983.
  18. *Guidelines for Chemical Transportation Risk Analysis*. Center for Chemical Process Safety, American Institute of Chemical Engineers, New York, 1995.
  19. Treichel, T. T., J. P. Hughes, C. P. L. Barkan, R. D. Sims, E. A. Phillips, and M. R. Saat. *Safety Performance of Tank Cars in Accidents: Probability of Lading Loss*. Report RA-05-02. RSI-AAR Railroad Tank Car Safety Research and Test Project. Association of American Railroads, Washington, D.C., 2006.
  20. Kawprasert, A., and C. P. L. Barkan. Effect of Train Speed on Risk Analysis of Transporting Hazardous Materials by Rail. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2159*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 59–68.
  21. Anderson, R. T., and C. P. L. Barkan. Railroad Accident Rates for Use in Transportation Risk Analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1863*, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 88–98.
  22. FRA. *Track Safety Standards Compliance Manual*. 2008. [http://www.fra.dot.gov/downloads/Safety/track\\_compliance\\_manual](http://www.fra.dot.gov/downloads/Safety/track_compliance_manual).
  23. Lin, S.-Y. *Optimizing of the Railroad Network Track Class*. Master's thesis. National Taiwan University, Taipei, Taiwan, 2010.
  24. *Analysis of Class I Railroads*. Policy and Economics Department, Association of American Railroads, Washington, D.C., 2007.
  25. Zarembski, A. M., R. Resor, and J. Cikota. *Technical Monograph: Estimating Maintenance Costs for Mixed High-Speed Passenger and Freight Rail Corridors*. Report. FRA, Washington, D.C., 2004.
  26. *GAMS: A User's Guide*. GAMS Development Corporation, Washington, D.C., 2008.
  27. Barkan, C. P. L., C. T. Dick, and R. Anderson. Railroad Derailment Factors Affecting Hazardous Materials Transportation Risk. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1825*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 64–74.

---

*The views expressed in this paper are those of the authors and do not necessarily reflect those of any of the acknowledged sponsors.*

*The Transportation of Hazardous Materials Committee peer-reviewed this paper.*

# Train Delay and Economic Impact of In-Service Failures of Railroad Rolling Stock

Bryan W. Schlake, Christopher P. L. Barkan, and J. Riley Edwards

Railcar condition directly affects the safety, the efficiency, and the reliability of freight railroad operations. Current railcar inspection practices are intended to identify defects before failure, but these practices generally do not enable preventive maintenance because manual, visual inspection is inherently limited. As a result, automated wayside condition-monitoring technologies have been developed to monitor rolling stock condition and facilitate predictive maintenance strategies. Improving the effectiveness of monitoring of railcar conditions could substantially reduce in-service failures and derailments, operational waste, and variability in rail operations and could enhance network productivity, capacity, and reliability. An analysis of the effect of lean production methods on main-line railway operations was conducted to determine the potential impact of improved railcar inspection and maintenance practices made possible by new, automated wayside technologies. Dispatch simulation software was used to quantify the magnitude and the variability of train delay as a function of both traffic level and severity of service outage. The results indicated that the annual cost caused by main-line delay was substantial compared with the annual cost of track and equipment damages from main-line derailments caused by mechanical causes. This work provided an analytical framework to assess the potential cost savings available through improved preventive maintenance strategies.

Since the early 1990s the U.S. railroad industry has made substantial investments in wayside detection systems capable of monitoring the condition of freight car components. From 1993 to 2008, U.S. Class I railroads spent over \$70 million on the development, installation, and maintenance of these systems (1). Previous economic analyses have justified these investments on the basis of the cost savings that resulted from a reduction in derailments. However, additional benefits associated with the reduction in main-line in-service failures (ISFs) should also be considered. An ISF occurs when a train stops on the main line because of a track or equipment defect. Although ISFs generally result in shorter, less costly delays than derailments, they occur much more frequently. U.S. Class I railroads experience thousands of equipment-caused ISFs per year; in contrast, the frequency of equipment-caused derailments ranges from 100 to 150 per

year (2). Thus, ISFs have a considerable financial impact caused by both the direct and indirect costs of main-line train delay.

To assess the economic impact of equipment-related ISFs, lean production principles were applied to railcar inspection and maintenance practices to determine the potential for both direct and indirect cost savings. Direct cost savings were addressed through an analysis of train delay, and indirect savings were addressed by analyzing the variability associated with train delay.

## AUTOMATED CONDITION-MONITORING TECHNOLOGY

Manual railcar inspections vary in their efficiency and effectiveness depending on inspection conditions and the experience or ability of individual car inspectors. A train may be inspected to the best ability of a particular car inspector, yet defects may be missed. Recognizing the inefficiency and subjectivity of manual inspection, the railroad industry has developed technologies to augment the efforts of human inspectors. These systems use various sensing mechanisms to measure heat, force, sound, and visual parameters in order to monitor the condition of railcar components. The technical maturity of these systems ranges from well established and in wide use by the industry for decades to systems that are still under development.

The first wayside detection systems were designed to identify defective components on passing trains in order to prevent derailments. Developed from the 1930s through the 1950s, these early technologies (e.g., hot bearing detectors and dragging-equipment detectors) provided a reactive means of defect detection, requiring a train to stop on the main line if a serious defect was identified (3–8). Although still widely used and effective in preventing derailments, these systems only provide component defect information shortly before or even after failure has occurred. Consequently, these technologies result in thousands of main-line ISFs each year, during which a train must either stop on the main line for a component to be repaired or have a railcar set out at a nearby yard or siding. In addition, because of the short latency period between condition detection and failure occurrence, these systems must be installed at frequent intervals across the railroad network, resulting in high installation and maintenance costs. Because of the high cost and limited predictive ability of these reactive technologies, railroads have sought the development of railcar condition-monitoring systems.

Automated condition-monitoring technology (ACMT) includes wayside detection systems capable of monitoring the condition of railcar components over time in order to facilitate preventive maintenance. Examples of ACMT include wheel impact load detectors, truck performance detectors, acoustic bearing detectors, hot wheel

---

Railroad Engineering Program, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 North Mathews Avenue, Urbana, IL 61801. Current affiliation for B. W. Schlake: Norfolk Southern Corporation, Mechanical Department, 110 Franklin Road, SE, Roanoke, VA 24042. Corresponding author: B. W. Schlake, bryan.schlake@nscorp.com.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 124–133.  
DOI: 10.3141/2261-14

detectors, cold wheel detectors, truck hunting detectors, wheel profile detectors, and machine vision. Detailed surveys of current way-side inspection technologies have been conducted by Lagnebäck (8), Steets and Tse (9), Bladon (10), Barke and Chiu (11), Robeda and Kalay (12), and Brickle et al. (13). The goal of ACMT is to facilitate condition-based maintenance, which is a “form of preventive maintenance based on vehicle performance and/or parameter monitoring” (8). Through the implementation of condition-based maintenance the physical condition of railcar components is monitored over time, allowing for trending analysis, early detection of deteriorating components, and the ability to predict component life. For example, through the use of acoustic bearing detectors, defective roller bearings can be identified much earlier than would have been possible with hot bearing detectors. By identifying railcar defects at an early stage and performing maintenance before component failure, railroads can reduce the likelihood of equipment-caused derailments and ISFs and take advantage of lower-cost, predictive maintenance strategies.

## LEAN RAILROADING

In 1990, the term “lean manufacturing” was introduced in a study at Massachusetts Institute of Technology (MIT). That study found Toyota’s production techniques to be superior to their competitors’ techniques in the automotive manufacturing industry (14). These findings helped to stimulate the use of lean methodology in other industries, and numerous companies throughout the world have since adopted lean methods. In the early 2000s, lean production techniques were formally applied to the North American railroad industry, although several key principles of lean railroading can be found in the earlier work of Sussman, Martland, and colleagues at MIT in the 1970s and 1980s (15, 16).

Lean is defined as the production of goods or services using minimal buffering costs (17). Sources of excessive buffering include both direct waste and variability. Direct waste is lean terminology for operations that are unnecessary. Examples in the railroad setting include accidents, ISFs, injuries, car damage, and unnecessary motion or information collection (16). Most managers focus on reducing these forms of direct waste, but another source of waste is variability.

Variability is a fundamental source of waste because it necessitates buffering in the form of extra inventory, capacity, or time (17). Common sources of excessive inventory buffering include variability in the frequency of ISFs caused by equipment, the length or severity of ISFs, and the level of maintenance required for various ISFs. These buffers can take the form of reserve supplies of empty or loaded freight cars and freight car components. Variability in train arrivals and unexpected defects requiring maintenance result in excess capacity buffers in railyards that may include extra yard tracks, car inspectors, or repair personnel. Finally, variability in run times, inspection and repair times, or labor availability may be buffered by adding slack time in the train schedule. All of these buffers are a result of the uncertainty inherent to various processes in the railroad system, and they lead to unnecessary costs in the form of indirect waste. Through the application of ACMT and condition-based maintenance, both the direct waste and the variability created by ISFs can be reduced.

ISFs result in both primary and secondary train delay. Primary, or exogenous delay, is direct delay caused by an external event affecting only the train experiencing an equipment defect (18). This includes either the time needed to repair a broken or defective railcar while on the main line or the time required to set out a car on a pass-

ing siding for future repair. Secondary delay, also called reactionary delay (19), is the delay to all other trains in the network affected by the service outage in some manner. According to lean principles, both primary and secondary delays that result from equipment-caused ISFs represent direct waste in the railroad network.

## ANALYSIS OF TRAIN DELAY

As U.S. freight traffic volumes have risen in recent years, research has been undertaken to better understand train delay and the impact of delay on network capacity and reliability. Schafer developed a train delay cost calculator to estimate the amount of train delay and the corresponding costs of broken rail-related derailments and service failures (20). In 2009, a study was conducted to understand the impact of higher train speed on freight railroad main-line capacity (21). More recent research has used Rail Traffic Controller (RTC) from Berkeley Simulation Software to analyze the impact of train type heterogeneity on railway capacity (22, 23). RTC is a dispatch simulation software package used throughout the North American railroad industry to simulate both freight and passenger operations (24, 25). Dingler provides an in-depth study of the use of RTC to investigate train delay and its relationship to capacity (23).

In the current study, RTC simulations were conducted on both single- and double-track routes, and the effect of train delay on unit coal traffic was analyzed. Results from these simulations were used to assess the potential economic impact of reducing equipment-related ISFs through the use of ACMT. In 2008, coal traffic comprised the greatest portion of tonnage and carloads originated (45% and 26%, respectively) and the most gross revenue (24%) among all commodities transported by U.S. Class I railroads (26). Additionally, unit coal trains operate on some of the highest-density rail corridors in North America, where delays caused by ISFs have the greatest impact. Another reason for analyzing unit coal traffic is the small variation in the design of railcars used to transport coal. This uniformity is advantageous for the design and implementation of machine vision and other condition-monitoring systems. For both of these reasons, the first applications for many ACMT systems will be unit train inspection on high-density coal routes.

### Dispatch Simulation Analysis: Single-Track Route

A common method of calculating train delay is through the use of dispatch simulation software, such as RTC. Previous rail capacity research using RTC provides a substantial background for the application of this software (23). A similar methodology was used in the current study, including the development of a representative North American single-track main-line subdivision and corresponding train schedule (Table 1).

The attributes used for this simulation were intentionally idealized for simplicity, as the purpose of this study was to determine the relative impacts of ISF length and traffic volume on main-line train delay. This analysis provides a baseline cost estimate for a typical rail route; however, estimation of costs for a particular route would require the actual characteristics specific to that route. Since route and traffic characteristics vary widely both among different railroads and within different subdivisions of an individual railroad, using actual characteristics would be highly specific, and results could not be universally applied. The analysis presented here is

TABLE 1 Parameters for Single-Track RTC Simulations

Route Characteristic	Train Characteristic
260 mi	Unit coal trains
10 mi between control points	115 cars, 6,325 ft long
8,000-ft signaled sidings	16,445 tons per train (loaded)
2.5-mi signal spacing	3,795 tons per train (empty)
Three-block, four-aspect signaling	0.78 hp/trailing ton
0% grade and curvature	Three SD70 4,300-hp locomotives
	Maximum speed: 50 mph

intended as a guide to the methodology and to provide insights regarding the relative magnitude of the effects that might be expected in a route-specific analysis.

The simulated traffic consisted of 115-car unit coal trains on a 260-mi, single-track route. To replicate typical coal route operations, loaded trains were run in one direction along the route, while empty trains were run in the opposite direction. During the simulation, a train was stopped at random times on the main line in order to replicate 1-, 3-, and 5-h ISFs. Twenty-four simulations were performed using a random variation of train starts, with each train departing the terminal within  $\pm 15$  min of its scheduled departure time. Total delay time was determined by subtracting the inherent delay for the base case (i.e., simulations without a service failure) from the delay for the simulations in which an ISF was initiated (Figure 1).

For each ISF length, average delay time increased exponentially with traffic volume. This exponential relationship is consistent with findings from other railway capacity research (18, 19). In addition, train delay curves increased more sharply for ISFs of greater length, indicating that main-line capacity on this route is more sensitive to longer ISFs.

Increases in train delay are mainly caused by increases in secondary delay. RTC simulations capture the shockwave effect created by an ISF, which is similar to those observed in highway traffic streams (27). Thus, a primary advantage of using RTC rather than a linear train delay cost calculator is the ability to more accurately predict delays to oncoming trains that are far away from the location of the ISF. In addition, RTC incorporates the time needed for braking and acceleration, thus providing more realistic train delay estimates.

### Dispatch Simulation Analysis: Double-Track Route

Since many of the corridors used to transport coal traffic experience high volumes and require a high level of maintenance, these routes are often built with multiple main-line tracks. As a result, RTC simulations were conducted using a double-track route. As before, 1-, 3-, and 5-h ISFs were randomly initiated to quantify total train delay at various traffic volumes. Twenty-two random simulations were conducted on the new route using the same train and route characteristics as before, except that the entire route contained two main-line tracks instead of a single track with sidings. During simulations, empty and loaded coal trains traveled in opposite directions, with each train type primarily using one specified track. As needed, trains were able to cross over to the other main-line track to run around stopped trains, following standard RTC dispatch simulation rules. Similar to the single-track simulations, train delay increased with an increase in either traffic volume or length of ISF (Figure 2).

As before, the increase in average delay is caused by increased secondary delay. Below a traffic volume of approximately 48 trains per day (i.e., one train on each main-line track per hour), average train delay increased linearly with traffic volume. However, above 48 trains per day, train delay increased exponentially.

In addition, variation in average delay times increased both at higher traffic volumes and for longer ISFs. A 95% confidence inter-

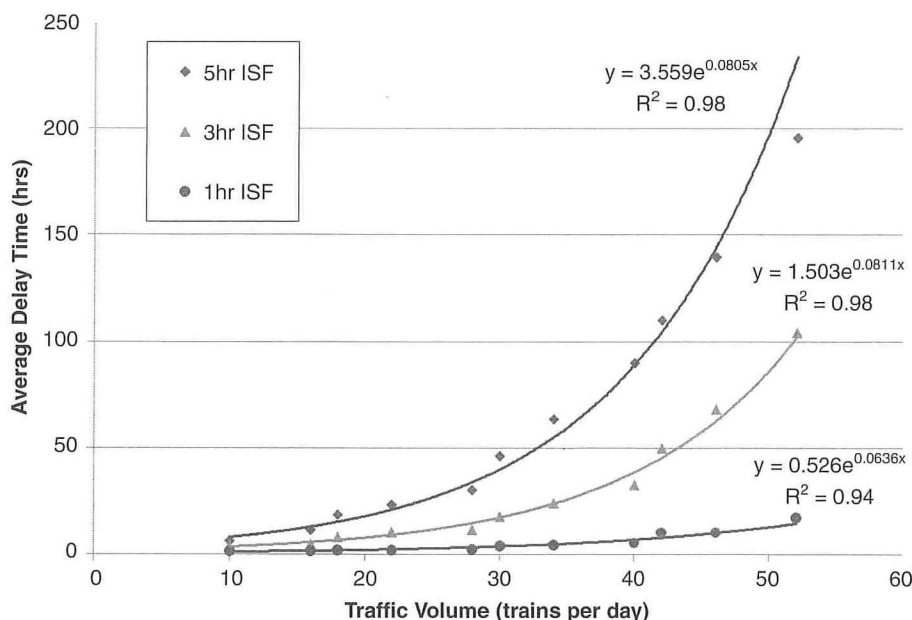


FIGURE 1 Average train delay data generated from RTC dispatch simulation software for varying ISF durations on single-track route.

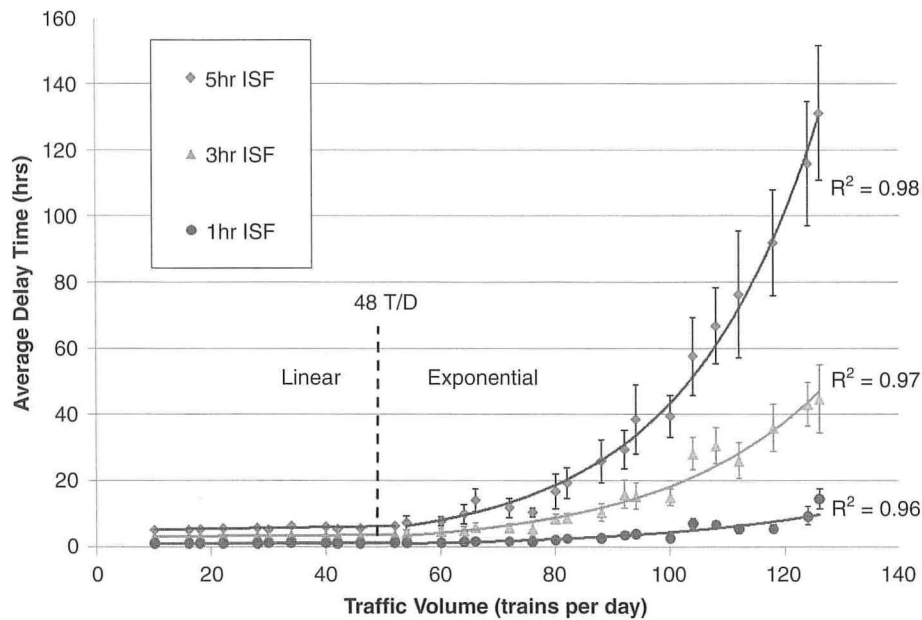


FIGURE 2 Total train delay as function of traffic volume with 95% confidence intervals for 1-, 3-, and 5-h ISFs on double-track route (T/D = trains per day).

val was used to determine the upper and lower bounds (denoted by the bars above and below each data point for average train delay in Figure 2). As traffic volume increased, the confidence intervals became larger, indicating an increase in variability. This variability in the system generates indirect waste that should be accounted for in addition to the direct waste created by train delay. An analysis of train delay variability is provided below.

### Estimation of Delay Cost for In-Service Failures

To determine the ISF-related delay cost incurred by a railroad, delay time is multiplied by a constant delay cost figure that includes four components: car cost, locomotive cost, fuel cost, and crew labor cost. The delay cost incorporates both the actual consumption of railroad company resources as well as the opportunity cost (in the case of cars and locomotives) of resources that are underutilized. A recent estimation of average total train delay cost was approximately \$213 per train hour for U.S. Class I railroads (20). This figure assumes an average of 69.2 cars per train and 2.7 locomotives per train.

The present study assumed 115 cars and three locomotives per train. Accounting for these changes, the total train delay cost increased to \$232 per train hour. This is assumed to be a conservative estimate, as it does not incorporate the lost revenue, or opportunity cost, caused by lading delay. Multiplying the constant hourly train delay value by average train delay times for various lengths of ISF resulted in cost curves following the same trends shown in Figure 1. Thus, for a 1-h ISF, the estimated total delay cost for the highest simulated traffic volume (52 trains per day) was approximately \$4,150 (17.9 train hours  $\times$  \$232 per train hour). Assuming a 5-h ISF, as traffic volume increased from 20 trains per day (approximately 74 annual million gross tons [ANMGT]) to 52 trains per day (approximately 192 ANMGT), delay costs increased from approximately \$5,000 to over \$45,000.

In most cases, ISFs result in delays shorter than 5 h. In general, less severe ISFs result in one of two scenarios: (a) the train will be inspected and repaired along the line-of-road and will continue service after repairs have been made, or (b) the defective railcar(s) is removed from the train and set out at a nearby storage track, passing siding, or yard. For either case, industry surveys estimate that the train will typically be delayed between 1 and 2 h. Using an approximation of 1.5 h, the potential costs of different types of ISFs at various traffic levels can be estimated as low (10 trains per day, or 37 ANMGT), medium (25 trains per day, or 92 ANMGT), and high (40 trains per day, or 148 ANMGT). According to RTC simulation data, a 1.5-h ISF will result in delay costs of approximately \$460, \$980, and \$2,850 for low-, medium-, and high-traffic routes, respectively.

### Cost Estimation Using Both Single- and Double-Track Routes

Main-line capacity is directly related to the physical infrastructure along a fixed route length. Increasing the number of tracks on a line (e.g., upgrading from single track to double track) results in a disproportionately greater increase in capacity (18). As a result, for traffic volumes under 48 trains per day (approximately 177 ANMGT), the delay costs caused by ISFs on a double-track route were almost negligible compared with those on a single-track route (Figure 3). The delay costs at these volumes were all less than \$1,500 regardless of ISF length. Although not shown in Figure 3, these costs displayed a linear trend, as seen with the delay time data (Figure 2).

A survey of U.S. Class I railroads indicated that they experience over 23,000 equipment-caused ISFs per year. These include ISFs caused by failed freight car components as well as those that occur in response to reactive wayside detectors, including hot bearing detectors and wheel impact load detectors. Total delay costs were

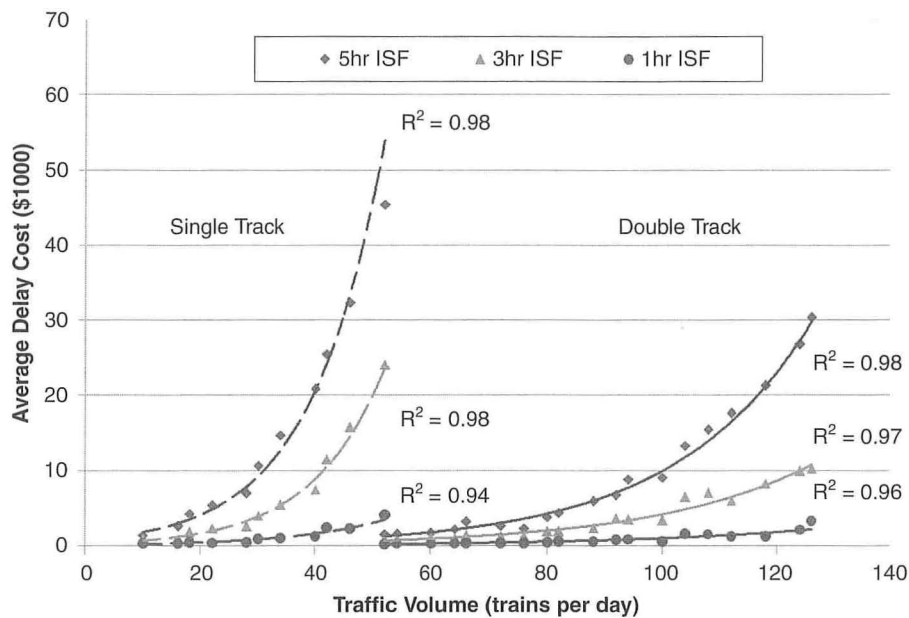


FIGURE 3 Total train delay as function of traffic volume for 1-, 3-, and 5-h ISFs on single- and double-track routes.

calculated on the basis of these numbers by using delay cost estimates for a sample 1.5-h ISF for different track types with average traffic volumes on the U.S. railroad network (Table 2).

Averages for the percentage of ton-miles on various single- and double-track routes were determined by using data from the *National Transportation Atlas* of the Bureau of Transportation Statistics (28). Although these ton-mile percentages include other types of traffic besides unit coal trains, final cost estimates should still be conservative because lading delay costs for other commodities will generally be higher. These data indicate that approximately 1% of U.S. ton-miles are transported on routes with more than two main tracks, but since delay costs would be minimal, these routes were not included. Multiplying delay costs by ton-mile percentages at each level of traffic, the annual failure cost is approximately \$15.2 million per year for U.S. class I railroads. For comparison, the average annual cost of track and equipment damages caused by main-line derailments was approximately \$35 million for the four largest U.S. Class I railroads

from 1999 to 2008 (2, 29). Thus, delay costs caused by main-line delay from ISFs appear to be substantial compared with the costs associated with derailments.

By identifying the costs associated with equipment-caused ISFs, railroads can more accurately assess the value of ACMT installation. Improved understanding of the cost benefits of ACMT can allow railroads to make better decisions regarding their technology implementation strategies. For example, since ISFs create much more train delay on single-track than double-track routes, railroads may choose to install a greater proportion of their ACMT in regions with large amounts of single track. However, it should not be assumed that increased implementation of ACMT will immediately result in proportionate reductions in equipment-caused derailments or ISFs. In some cases, new ACMT installations could initially result in additional ISFs caused by false alarms. However, as these technologies are further tested and improved, appropriate component inspection thresholds will be determined, and system accuracies will increase. A key area of future research will be the development of a condition-monitoring efficiency metric to assess the proportion of equipment-caused failure costs that can be recovered using ACMT.

TABLE 2 Train Delay Cost and Percentage of Ton-Miles by Track Type Based on RTC Simulation Data for 1.5-h ISF with Various Traffic Volumes (28)

Average ANMGT	Single Track		Double Track	
	Delay Cost (\$)	Ton-Miles (%)	Delay Cost (\$)	Ton-Miles (%)
<40 (~37)	460	31.3	350	0.9
40–60	590	17.9	350	2.3
60–100	1,000	18.7	360	6.2
>100 (~110)	2,170	5.5	440	16.3
Total	556.45 <sup>a</sup>	73.5	104.83 <sup>a</sup>	25.6

<sup>a</sup>Weight average based on percentage of ton-miles.

## VARIABILITY IN TRAIN DELAY

Because variability results in necessary buffering (e.g., added slack time in train schedules), it is a fundamental source of waste (16). RTC simulations indicated that the variability associated with equipment-related main-line train delay increased both with traffic volume and ISF length. Both factors were analyzed separately by using frequency diagrams with the number of trains delayed versus the length of individual train delay (in minutes) per 100 train miles. The lengths of individual train delays were divided into 10 individual frequency bins. These frequency bins ranged from 0 to over 270 min for the single-track route and from 0 to 90 min for the double-track route.

Trains in the first bin experienced little or no delay, while the trains in the last bin incurred the largest amounts of delay. Curves showing the percentage of total trains are also provided to more clearly illustrate the increased variability.

### Length of In-Service Failure and Variability

The first factor analyzed was the length of ISF. To best evaluate the impact of this factor, simulations using the single-track route were chosen at the highest simulated traffic volume (52 trains per day, or 192 ANMGT). Although this traffic volume is higher than that of most single-track operations, using exaggerated conditions in the simulations highlights the impact of ISF length on main-line capacity, allowing qualitative analysis of this factor. In the previous sections, total delay was calculated by subtracting the delay for the base case from the delay for 1-, 3-, and 5-h ISFs. In this analysis, the frequency of delay for ISFs was compared with the base case (0-h ISF) to show changes in train delay variability. Therefore, train delay for this analysis was defined as the difference between the minimum, or unopposed, run time and the actual time it takes a train to traverse the route (22). Given this definition, over 60% of the trains in the base case experienced between 30 and 60 min of delay (Figure 4).

As the length of ISF increased from 0 to 5 h, both average train delay and train delay variability increased. For the various ISF lengths, the distribution of the data shifted from a skewed distribution to one that was more symmetrical, with the modal value increasing (i.e., shifting to the right) as the length of delay increased. For the base case, most trains experienced 30 to 60 min of delay (per 100 train miles), while for a 5-h ISF, the modal value was 120 to 150 min. The distribution curves became wider and shorter for longer ISFs, indicating an increase in variance. This increase in variance affects a railroad's level of service, because when a higher percentage of trains is delayed, more customers are affected, resulting in greater costs to the railroad.

For double-track routes, the increases in average delay and variability were not as evident. The double-track route described above was analyzed using the highest simulated traffic volume (126 trains per day, or 465 ANMGT). Since there is less total delay on the double-

track route, 10-min frequency bins were used instead of the 30-min bins shown in Figure 4. The data show that for a 1-h ISF, 90% of the trains experienced little or no effect (i.e., 0 to 10 min of delay), and none of the trains were delayed for more than 30 min (Figure 5a). However, for a 5-h ISF, only 46% of the trains experienced little or no effect, and over 25% of the trains experienced a delay greater than 30 min. As expected, the longer the ISF, the greater the number of trains affected.

The distribution of the frequency bars provides information about the variance of the train delays. To more clearly see these distributions, the data are displayed for only the delays greater than 10 min (Figure 5b). For 1- and 3-h ISFs, the frequency of delays decreased as the length of delay increased. However, the 5-h ISF case followed a different trend. Unlike the other distributions, the distribution of the 5-h ISF increased at the highest levels of individual train delay. These data suggest that when a service outage exceeds a certain threshold, the shockwave affecting the network becomes larger and more unpredictable. Thus, although the effects were not as pronounced as in the single-track route, the length of service outage affected the amount of operational waste generated in the network.

### Traffic Volume and Variability

For the second factor, traffic volume, data were analyzed for a 5-h ISF on both the single- and double-track routes with traffic volumes set from 16 to 52 and 64 to 126 trains per day, respectively (Figures 6 and 7). Similar to the impact of ISF length, increased traffic volume resulted in increased variability in train delay. As before, the distribution curves for the single-track data shifted to the right and became wider and shorter as traffic volume increased (Figure 6).

The impact of traffic volume on train delay was less apparent on double-track routes because of the much greater line capacity (Figure 7). However, it is clear that average train delay and train delay variability increased at higher traffic volumes. For very high traffic volumes (i.e., above 82 trains per day, or 302 ANMGT), the train frequency distributions were not normally distributed (Figure 7b). Instead, the number of trains delayed decreased until the length of individual train delay reached approximately 70 min and then began

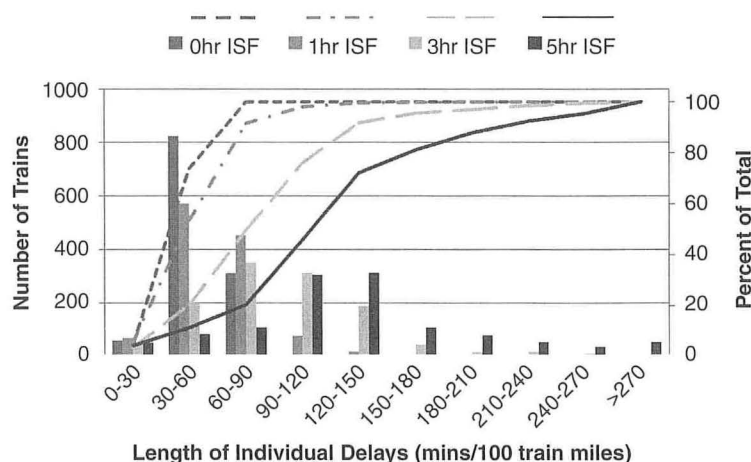


FIGURE 4 Frequency diagram showing amount of delay for each train caused by various ISFs on single-track route with 52 trains per day.

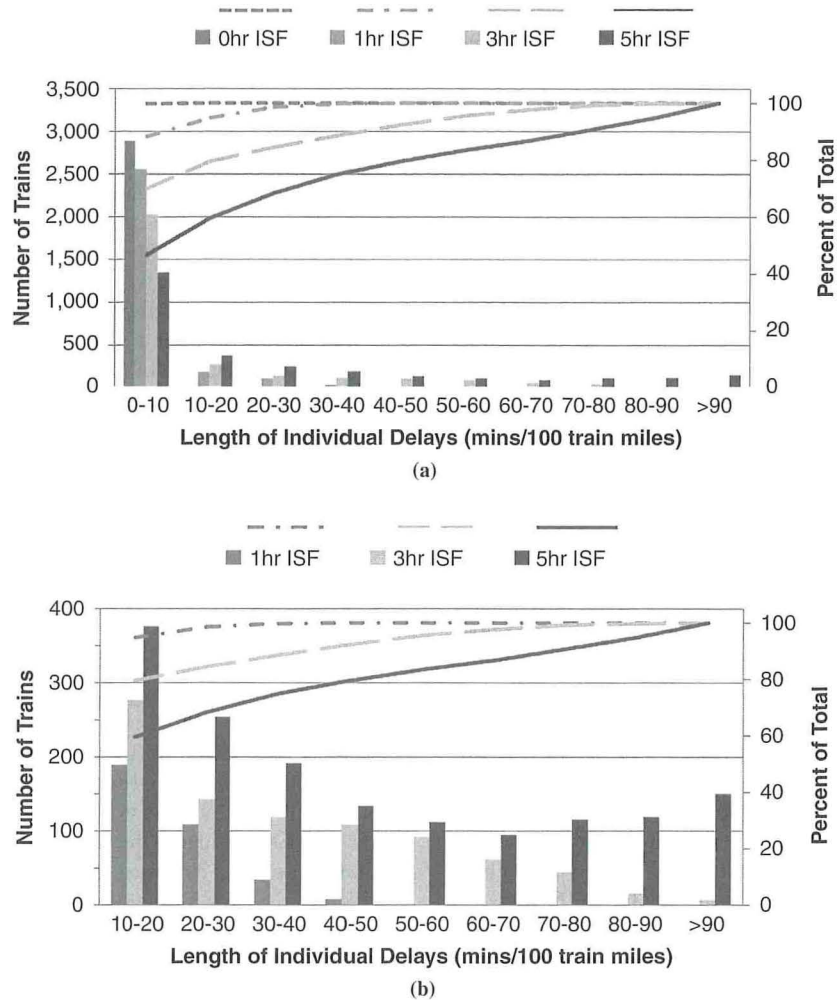


FIGURE 5 Frequency diagrams showing amount of delay caused by various ISFs on double-track route with 126 trains per day for (a) each train and (b) trains experiencing delays longer than 10 min.

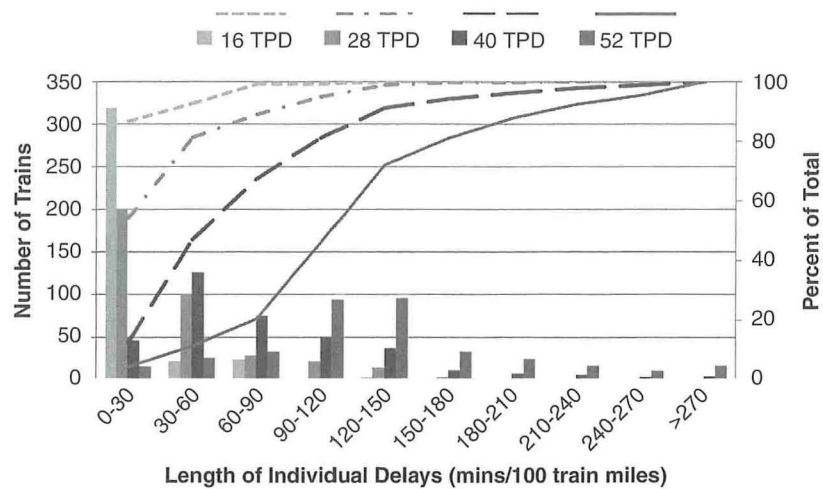


FIGURE 6 Frequency diagram showing amount of delay for each train caused by 5-h ISF on single-track route with varying traffic volumes (TPD = trains per day).

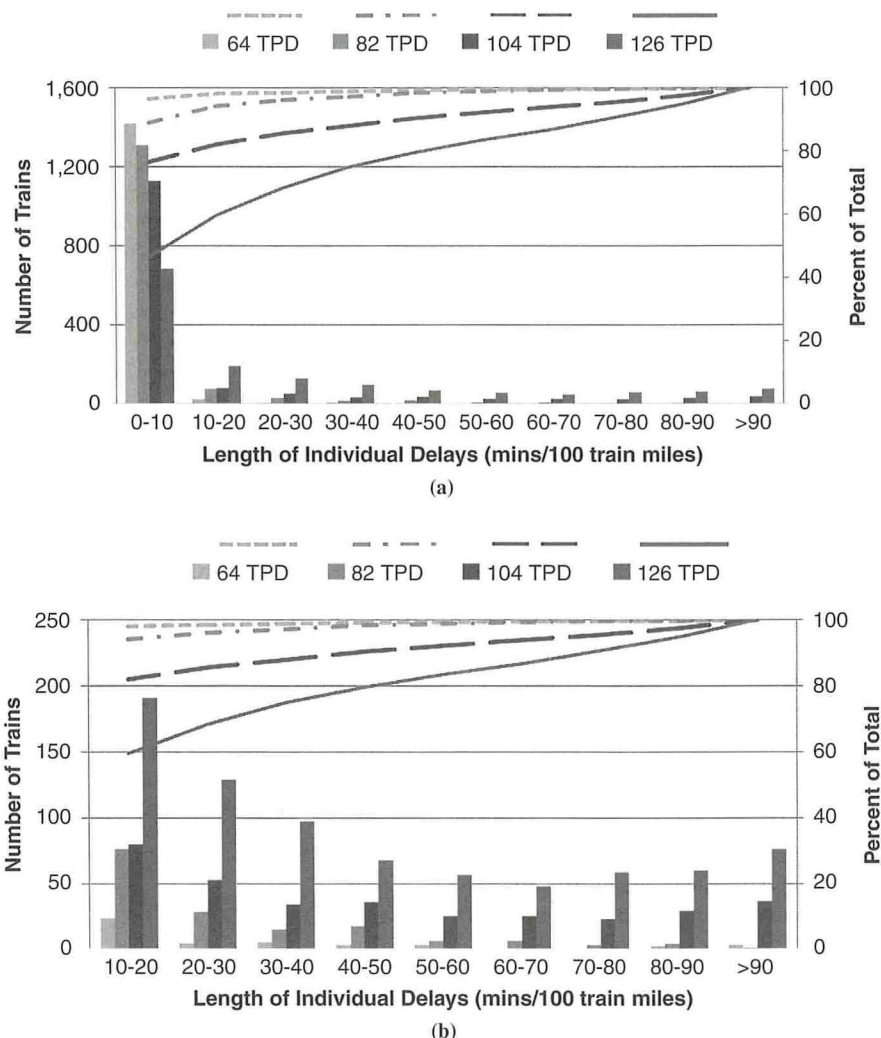


FIGURE 7 Frequency diagrams showing amount of train delay caused by 5-h ISF on double-track route with varying traffic volumes for (a) each train and (b) trains experiencing delays longer than 10 min.

to increase. In other words, the mechanical reliability of freight cars becomes more important at higher traffic levels. Although this is true for double-track routes, its impact is much greater on single-track routes.

## CONCLUSIONS

The effectiveness of railcar condition monitoring has a large impact on rail transportation efficiency. Dispatch simulation software was used to analyze the effect of ISF duration and traffic volume on single- and double-track versions of a hypothetical route to estimate train delay. The simulations indicated that both traffic volume and ISF length had a nonlinear effect on delay, with traffic volume having an exponential effect. The associated costs may be higher than previously estimated, especially for high traffic volumes. Based on RTC simulation data, the estimated cost of the direct waste caused by these main-line delays is approximately \$15.2 million per year for U.S. Class I railroads. Although train delay costs caused by ISFs are not often considered in economic analyses of ACMT systems, they

are substantial compared with the track and equipment damages associated with derailments. Another factor not often considered is the large variability in train delay at high traffic volumes. When variability increases, there is a higher probability that more trains will experience longer delays, resulting in indirect waste in the form of increased time buffers. Although the costs caused by variability are more difficult to quantify, this negatively affects the level of service that railroads can offer their customers. Additional failure costs associated with train delay can be recovered by improving railcar inspection and maintenance practices and reducing the likelihood of equipment-caused main-line ISFs. These costs are considered in more detail by Schlake, who provides a thorough study of the impact of ACMT on freight railroad safety and efficiency (30).

## FURTHER RESEARCH

The present study developed a framework to assess the potential impact of equipment-related ISFs on railroad main-line efficiency. In order to determine the total costs and benefits of implementing

ACMT, a metric should be developed to determine the effectiveness with which critical railcar components can be monitored. This metric would be a function of both the accuracy of ACMT systems and the statistical probability that a specific railcar defect would cause an ISF. In order to develop this metric, appropriate data must be collected and analyzed from field installations of ACMT and from records of ISF occurrences. These analyses would allow the proportion of ISF costs that could be recovered using ACMT to be determined. In addition, by determining the most critical railcar components (i.e., those that are most likely to cause an ISF or derailment), automated inspection efforts could be directed most effectively to prevent the maximum number of ISFs and reduce the variability in their duration. Once this metric is developed for unit coal train operations, these methods could also be applied to other traffic types, such as intermodal operations.

Findings from this study indicate that ISFs have a much greater impact on single-track than on double-track operations. Both train delay costs and delay variability were higher for single-track routes. As a result, an important area of future research is determining optimal locations for ACMT installations while taking the cost of ISF-caused main-line delays into account. Ouyang et al. found that simply installing ACMT at the busiest locations on a network yielded suboptimal results (31). However, it may be beneficial to develop additional optimization models aimed at minimizing the cost of main-line train delay. This would allow railroads to not only ensure full coverage of their railcar fleets by ACMT, but also provide the capability to prevent equipment-caused ISFs in areas of particularly high traffic or limited capacity.

Other technologies may also provide additional means of reducing equipment-caused ISFs. Technologies capable of performing onboard diagnostics for railcar components could result in substantial reductions in train delays. Electronically controlled pneumatic brakes are a prime example of this type of technology. As with ACMT, a metric could be developed to determine the proportion of equipment-caused ISFs that could be reduced through the use of electronically controlled pneumatic brakes. Although it is not likely that the reduction in main-line train delays would justify the cost of retrofitting a fleet of cars with electronically controlled pneumatic brakes, this is a factor that should be considered in cost-benefit analyses for this technology.

Another area of future research is the quantification of train delay costs associated with major derailments. For longer main-line service outages (e.g., 24 to 48 h), in the present study performing RTC simulations was impractical because of the amount of time and computational resources required. In addition, because of the rerouting that occurs and other complexities that accompany derailments, actual train delay costs cannot be easily estimated using either linear or exponential estimation methods. Since these costs would vary substantially among different routes, the best approach may be to use empirical analysis with historical derailment data.

Finally, future research should assess the costs associated with ACMT implementation, including purchase and installation, system maintenance, electronic infrastructure, integration (including the time and expense of incorporating institutional changes), and additional railcar maintenance resulting from improved inspection effectiveness. A thorough analysis of both the costs and benefits of improved railcar condition monitoring will provide railroad management with the tools necessary to make informed decisions regarding the implementation of ACMT.

## ACKNOWLEDGMENTS

The authors thank Eric Wilson and others of Berkeley Simulation Software for the use of Rail Traffic Controller (RTC). Mark Dangler of the University of Illinois at Urbana-Champaign (UIUC) provided technical insight regarding RTC and train delay analyses. The authors thank Tom Guins of the Association of American Railroads for his assistance on this project. In addition, the authors thank Hayden Newell and Nate Stoehr of Norfolk Southern, Steve Beckwith of Union Pacific Railway, Don Lauro of CSX, and Darrel Iler of CN for providing train delay data and technical insight. Bryan Schlake has been supported in part by a CN Railroad Engineering Research Fellowship at UIUC. J. Riley Edwards has been supported in part by grants to the UIUC Railroad Engineering Program from CN, CSX, Hanson Professional Services, Norfolk Southern, and the George Krambles Transportation Scholarship Fund.

## REFERENCES

1. Robert, W., A. Aeppli, and P. Little. *Post-Audit of Wayside Detector Costs and Benefits*. Cambridge Systematics Inc., Cambridge, Mass., 2009.
2. Office of Safety Analysis, FRA, U.S. Department of Transportation. Sec. 3.01-3.04, 9.01, and 9.05. [http://safetydata.fra.dot.gov/officeof\\_safety/](http://safetydata.fra.dot.gov/officeof_safety/). Accessed Dec. 14, 2009.
3. Post, W. M. Protective System for Railways. U.S. Patent 2063336, 1936.
4. Post, W. M. Protective System for Railways. U.S. Patent 2095616, 1937.
5. Burpee, C. M. (ed.). *Railway Engineering and Maintenance Cyclopedia*, 6th ed. Simmons-Boardman Publishing Company, Chicago, Ill., 1945.
6. Austin, K. B. Hotbox Signal for Railway Trains. U.S. Patent 2486546, 1949.
7. Gallagher, C. A., and W. M. Pelino. Hot-Box Detector. U.S. Patent 2880309, 1959.
8. Lagnebäck, R. *Evaluation of Wayside Condition Monitoring Technologies for Condition-Based Maintenance of Railway Vehicles*. Licentiate thesis. Luleå University of Technology, Luleå, Sweden, 2007.
9. Steets, P. G., and Y. H. Tse. Conrail's Integrated Automated Wayside Inspection. *Proc., IEEE/ASME Joint Railroad Conference 1998*, Piscataway, N.J., 1998, pp. 113-125.
10. Bladon, T. Predictive Condition Monitoring of Railway Rolling Stock. *Proc., Conference on Railway Engineering*, Darwin, Australia, 2004.
11. Barke, D., and W. K. Chiu. Structural Health Monitoring in the Railway Industry: A Review. *Structural Health Monitoring*, Vol. 4, 2005, pp. 81-93.
12. Robeda, J., and S. Kalay. Technology Drives U.S. Train Inspections. *International Railway Journal*, Vol. 48, No. 5, 2008, pp. 47-50.
13. Brickle, B., R. Morgan, E. Smith, J. Brosseau, and C. Pinney. *Wheelset Condition Monitoring*. RSSB report for Task T607. TTCL, Ltd., London, 2008.
14. Womack, J. P., D. T. Jones, and D. Roos. *The Machine That Changed the World: The Story of Lean Production*. HarperCollins Publishers, New York, 1990.
15. Sussman, J. *Introduction to Transportation Systems*. Artech House, Inc., Norwood, Mass., 2000.
16. Hopp, W. J., and M. L. Spearman. To Pull or Not to Pull, What Is the Question? *Manufacturing and Service Operations Management*, Vol. 6, No. 2, 2004, pp. 133-148. [www.factoryphysics.com](http://www.factoryphysics.com). Accessed May 7, 2010.
17. Dirnberger, J. R., and C. P. L. Barkan. Lean Railroad for Improving Railroad Classification Terminal Performance: Bottleneck Management Methods. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1995, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 52-61.
18. Mattsson, L.-G. *Railway Capacity and Train Delay Relationships*. Springer, Berlin, 2007, pp. 129-150.

19. Gibson, S., G. Cooper, and B. Ball. Developments in Transport Policy: The Evolution of Capacity Charges on the U.K. Rail Network. *Journal of Transport Economics and Policy*, Vol. 36, 2002, pp. 341–354.
20. Schafer, D. H. *Effect of Train Length on Railroad Accidents and a Quantitative Analysis of Factors Affecting Broken Rails*. MS thesis. University of Illinois at Urbana–Champaign, 2006.
21. Harrod, S. Capacity Factors of a Mixed Speed Railway Network. *Transportation Research Part E*, Vol. 45, No. 5, 2009, pp. 830–841.
22. Dinger, M. H., Y.-C. Lai, and C. P. L. Barkan. Impact of Train Type Heterogeneity on Single-Track Railway Capacity. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2117, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 41–49.
23. Dinger, M. H. *Understanding the Impact of Operations and New Technologies on Railroad Capacity*. MS thesis. University of Illinois at Urbana–Champaign, 2010.
24. Parsons Brinckerhoff Quade & Douglas, Inc. *The Long-Term Financial Feasibility of the Northwestern Pacific Railroad*. Final report. 2002. <http://www.northcoastrailroad.org/feasibility.html>. Accessed March 15, 2009.
25. Washington Group International, Inc., LOSSAN Rail Corridor Agency, and IBI Group. *RTC Simulations: LOSSAN North Railroad Capacity and Performance Analysis*. 2007. [www.sbcag.org/Meetings/SCSPC/2007/February/Item5LOSSANSR.pdf](http://www.sbcag.org/Meetings/SCSPC/2007/February/Item5LOSSANSR.pdf). Accessed March 15, 2009.
26. *Railroad Facts*. Association of American Railroads, Washington, D.C., 2009.
27. Garber, N. J., and L. A. Hoel. *Traffic and Highway Engineering*, 4th ed. Cengage Learning, Toronto, Ontario, Canada, 2009.
28. Research and Innovative Technology Administration, Bureau of Transportation Statistics, U.S. Department of Transportation. *National Transportation Atlas Database*. 2006. [http://www.bts.gov/publications/national\\_transportation\\_atlas\\_database](http://www.bts.gov/publications/national_transportation_atlas_database). Accessed March 16, 2010.
29. Schlake, B. W., C. P. L. Barkan, and J. R. Edwards. Impact of Automated Inspection Technology on Unit Train Performance. *Proc., Joint Rail Conference 2010*, Urbana, Ill., 2010.
30. Schlake, B. *Impact of Automated Condition Monitoring Technologies on Railroad Safety and Efficiency*. MS thesis. University of Illinois at Urbana–Champaign, 2010.
31. Ouyang, Y., X. Li, Y.-C. Lai, C. P. L. Barkan, and A. Kawprasert. Optimal Locations of Railroad Wayside Defect Detection Installations. *Computer-Aided Civil and Infrastructure Engineering*, 2009, Vol. 24, pp. 1–11.

---

*The Railroad Operating Technologies Committee peer-reviewed this paper.*

# Basic Tool Kit for Estimation of Intermodal Rail Cost

Dan P. K. Seedah, Robert Harrison, and James R. Blaze

**Federal and state transportation planners and others seeking to analyze transportation systems find few publicly available rail analysis models to estimate the operational costs and environmental impacts of rail movements. Moreover, data to populate such models and to test public policy considerations for evaluating public-private partnerships are generally difficult to obtain. This paper, a product of a study funded by Region 6 of the University Transportation Center Program, offers stakeholders the building blocks to develop an integrated rail analysis model capable of testing railway operational and capital investment changes. The paper also reviews the current state of rail modeling, examines selected rail models, and presents the findings of a preliminary intermodal rail costing model developed in the work.**

Analyzing rail operational benefits and costs is an inherently complex process. Forkenbrock (1) and Bereskin (2) suggest several factors that may contribute to this complexity, including technological innovations, joint production among rail companies, lack of data on specific expenditures pertaining to individual freight movements, and economies of scale, scope, and density. Furthermore, the high capital costs required to construct and maintain rail service obscure the ability of outside analysts to determine how much it actually costs the railroad to transport any given shipment. Nevertheless, an understanding of and an ability to simulate rail operations are essential for public transportation stakeholders to effectively examine costs and environmental impacts when they are asked to partially fund a private railway project.

Methods to determine rail costs have always been central to rail operations. Since deregulation academicians and government organizations have developed models to examine various components of rail operations. Noted authors like Bereskin (2-4), Forkenbrock (1), Caves et al. (5, 6), Ivaldi and MacCullough (7), and Spady and Freidlaender (8) have reported on the railroad industry's achievement of productivity gains over time and through mergers and have provided evidence of the nonlinearity of many rail costs (3). The existence of economies of scope in the railroad industry and the production of different outputs at different cost levels have been documented as well as the potential of increases in rail traffic to cause diseconomies as a result of traffic delays under different conditions (2).

While economists such as Bereskin have developed highly refined econometric models of rail cost that include track capacity, government agencies such as the Surface Transportation Board (STB) are more limited in the types of tools they use in determining public policy and the economic impacts of rail service change or whether rates are in line with variable costs. For more than two decades, STB has used the uniform rail costing system (URCS) model, a general-purpose costing system for estimating variable and total unit costs for Class I U.S. railroads. While the URCS model has some limitations, it is still the official tool used by the STB.

URCS uses system average units based on cost relationships and system data for the large Class I railroads. Although STB updates the data annually, the basic structure of the model remains as it was when it was developed. For example, there is no clear way to specify costs for intermodal double-stacked containers.

Researchers have identified four primary problems with URCS. First, the model uses linear percent variable equations to allocate expenses to specific operating activities based on a cross-sectional regression of cost data against traffic data for the Class I railroads of the 1980s, using a several-year time series. The equations therefore do not account for recent industry changes (e.g., mergers, increasing size, and traffic carried) that affect the operational costs of railroads (3). Furthermore, the linear nature of the model is contrary to the finding that certain rail costs are nonlinear in nature. Second, URCS uses system averages based on annual operating expenses and traffic data reported by the Class I railroads (9). System averages may not reflect the actual railroad rates charged by carriers, geographical location, technological improvements, or system performance (10). To counter this limitation, URCS gives users the flexibility of substituting cost data developed by STB with user-generated costs. However, such data can be difficult to obtain. Third, URCS does not account for changes in fuel prices and rail traffic. The model does not have an input for fuel cost, which is a serious flaw since fuel price surges and surcharges for fuel are extremely volatile. Finally, URCS does not have the ability to estimate greenhouse gas emissions produced during rail operations, which is essential for comparison with other transport modes when performing public policy analysis.

Another promising rail costing model is the rail energy cost analysis package (RECAP II) developed in 1985 by Smith (11). RECAP II was built around the Association of American Railroads (AAR) train energy model and enhanced with the development of a driver program, a cost model, and a data matrix generated by the track maintenance cost model (11). However, RECAP II has not been updated since it was first developed. Because of the noted limitations of URCS and RECAP II, researchers at the Center for Transportation Research deemed it necessary to develop a new transparent rail operations costing model. The following section presents an integrated rail analysis modeling framework of three building blocks comprising external parameters, asset management, and operating parameters.

---

D. P. K. Seedah and R. Harrison, Center for Transportation Research, University of Texas at Austin, Suite 4.202, 1616 Guadalupe Street, Austin, TX 78701. J. R. Blaze, Zeta-Tech Associates Inc., 900 Kings Highway, N., Cherry Hill, NJ 08034. Corresponding author: D. P. K. Seedah, dseedah@mail.utexas.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 134-142.  
DOI: 10.3141/2261-15

The next long section describes a preliminary intermodal rail costing model called CTRail. The fourth section offers two case studies, and the fifth section provides a conclusion.

## RAIL ANALYSIS MODELING FRAMEWORK

### External Parameters

External parameters of rail costing include the influence of rail traffic and rail demand on individual rail movements. Hay notes that railroads incur overhead costs regardless of whether equipment is used (12). Overhead and direct costs are distributed over the volume of traffic handled. The greater the rail traffic, the lower the share of fixed cost borne by a single unit of traffic. However, as volumes increase, unit cost begins to increase again as congestion, delays, and maintenance costs increase. Should additional capacity be provided, unit cost begins to fall (12). The external parameters block assists stakeholders in measuring the impact of rail volumes on rail capacity and its effect on individual rail movements. This is important as demand drives the volume of traffic on the network at any given time. The external parameters block serves as an input for the operating parameters block.

### Asset Management

Rail asset management involves the management of railroad equipment and personnel. Items include equipment maintenance, asset depreciation, capital or interest charges, and overhead and personnel management. Equipment maintenance includes taking stock of the number of units of specific equipment and the cost associated with maintaining the equipment. Asset depreciation accounts for the reductions in value of owned equipment. Capital or interest charges are costs accrued from the purchase of new equipment and the upgrading or development of rail infrastructure such as tracks and signals. Overhead and personnel management comprise the salaries and benefits paid to railroad employees. Asset management may also include equipment leasing and rental. The asset management block also provides data to be used in the operating parameters block when simulating the costs associated with individual rail movements.

### Operating Parameters

Operating parameters involve the simulation of a single train through a prespecified set of inputs such as route characteristics, type of locomotive, type of rail cars, commodities transported, emission rates, crew wages, and loading and unloading operations. Some inputs of the operating parameters block, such as travel time and maintenance costs, are calculated from the two other building blocks. The external parameters block determines the delay of rail operations on the basis of capacity and demand, and the asset management block calculates the costs associated with equipment depreciation and track maintenance.

## INTERMODAL RAIL COSTING MODEL

As part of this study, a preliminary intermodal rail model (CTRail), which forms part of the line haul section of the operating parameters block, was developed. The core equations governing the line haul

model were adapted from work by DeSalvo (13), Hay (12), and Avallone et al. (14). DeSalvo investigated the various productivity relationships, isoquants, and returns to scale for the rail line haul process. Some of the equations used in his paper (mainly the resistance equations) have since been modified by Hay (12) and Avallone et al. (14).

CTRail enables stakeholders to measure operational differences between trailer-on-flat-car (TOFC) units and double-stacked containers on flat or well cars in intermodal service. It also allows for the calculation of gallons of fuel consumed, greenhouse gas emissions produced, the effect of operational differences when using multiple locomotives or car types, the influence of delay, and other route-specific characteristics such as grade changes and road curvature. This initial intermodal model is mechanistic in nature and uses as inputs various factors such as cargo weight, energy consumption, and expert estimates of maintenance and crew labor costs. CTRail is divided into eight costing or analysis modules (described below) that work together to provide cost estimates for line haul movement:

- Cargo weight, number of containers, and rail car configuration;
- Locomotive selection;
- Train-in-motion calculations;
- Fuel consumption and cost;
- Locomotive emissions;
- Crew labor cost;
- Maintenance cost; and
- Capital cost and investment cost.

### Cargo Weight, Number of Containers, and Rail Car Configuration Modules

More than 10 types of intermodal rail cars are used today, each with its own tare weight, cargo capacity, and load limit. CTRail allows users to select any of the available rail cars and container types. The weight of a single car ( $w_{ci}$ ) can be said to be equivalent to

$$w_{ci} = c_i + x_i + k_i \quad (1)$$

where

- $c_i$  = tare weight of the rail car,
- $x_i$  = tare weight of the container (if intermodal service), and
- $k_i$  = cargo weight.

For an intermodal double-stacked service, the weight of a single car ( $w_{si}$ ) is equivalent to

$$w_{si} = c_i + 2(x_i + k_i) \quad (2)$$

Given a certain number of cars ( $N_c$ ) or when simulating an intermodal TOFC service, the total weight of cargo ( $W_c$ ) will be

$$W_c = \sum_{i=1}^{N_c} w_{ci} \quad (3)$$

For an intermodal double-stacked service, given a certain number of containers ( $X$ ), the total number of cars will be

$$N_x = \frac{X}{2} \quad (4)$$

and the total weight of cargo ( $W_s$ ) will be

$$W_s = \sum_{i=1}^{N_s} w_{s_i} \quad (5)$$

### Locomotive Selection Module

The total number of locomotives is dependent on the horsepower of each locomotive and the desired horsepower per trailing ton (HPTT) ratio. The HPTT ratio is determined by railroads and varies by route and service type. It dictates the desired maximum speed of the train. The typical ratios used by Class I railroads vary from 2.5 to 3.5 HPTT for intermodal to less than 2.5 for coal and other heavier cargo. CTRail allows the user to specify either the desired ratio or maximum speed and then calculates the total horsepower required.

The total number of locomotives ( $N_L$ ) is calculated on the basis of the required horsepower divided by the specified horsepower of each locomotive ( $hp_{L_i}$ ):

$$N_L = \frac{hp_{\text{required}}}{hp_{L_i}} \quad (6)$$

Given the weight of a single locomotive as  $w_{L_i}$ , the total weight of all the locomotives is equal to  $W_L$ , and the total weight of the train can be calculated as  $W$ :

$$W_L = \sum_{i=1}^{N_L} w_{L_i} \quad (7)$$

$$W = W_c + W_s + W_L \quad (8)$$

### Train-in-Motion Module

According to Hay, train movement and speed are opposed by resistances that must be overcome by the propulsive force (also called tractive effort) of the locomotive (12). These forces contribute to rail operation and overall operating costs (12). Internal resistance of the locomotive, resistances varying directly at the axle loading (journal friction, rolling resistance, and track resistance), flange resistance, air resistance, and track modulus resistance are always present during train movement. An expression for these resistances, known as train resistance, was developed empirically. Wind resistance, external axle loading resistance, curve resistance, grade resistance, acceleration resistance, and inertia (starting) resistance are only present intermittently but are also estimated through empirical relationships (12). At present these resistances are not included in the train resistance equations. CTRail currently calculates train speed as a function of tractive effort, train resistance, curve resistance, and grade resistance.

#### Tractive Effort

Tractive effort, the force required to pull a train, is determined as

$$F_T = (hp_e - hp_a) \times 375 \times \frac{e}{V} \quad (9)$$

where

$F_T$  = tractive effort (lb),  
 $hp_e$  = engine shaft horsepower,

$hp_a$  = horsepower of the auxiliaries,  
 $e$  = efficiency of the locomotive, and  
 $V$  = speed (mph).

The most common interpretation (12, 13) for Equation 9 is Equation 10, where  $e$  is 0.82 and  $hp$  is the manufacturer's rated horsepower:

$$F_T = \frac{300 \text{ hp}}{V} \quad (10)$$

#### Train Resistance

Train resistance is modeled using the basic Davis equation, the modified Davis equation, and the adjusted Davis equation. The basic Davis equation is known to result in resistances higher than the modified and adjusted versions but is still relevant for calculating drag and flange friction resistance for locomotives. Using the basic Davis equation, the train resistance for a single locomotive ( $R_i$ ) is

$$R_i = 1.3w_i + 29a_i + bw_iV + cZV^2 \quad (11)$$

where

$w_i$  = weight of a single locomotive,  
 $a_i$  = number of locomotives axles,  
 $V$  = train speed,  
 $Z$  = locomotive cross-sectional area (120 ft<sup>2</sup>),  
 $b$  = coefficient of flange friction (0.03 for locomotives), and  
 $c$  = drag coefficient of air (0.0025 for locomotives).

With the substitution of the values of  $b$ ,  $c$ , and  $Z$ , the total train resistance of all locomotives is  $R_L$ , where  $W_L$  is the total weight of all locomotives,  $A_L$  is the total number of axles of all locomotives, and  $N_L$  is the number of locomotives:

$$R_L = 1.3W_L + 29A_L + bW_LV + cN_LV^2 \quad (12)$$

Improvements in railroad operations resulted in the need to adjust the basic Davis equation, especially for rail cars (12). The modified Davis equation is similar to AAR's equations and is appropriate for relatively high weights of 70 tons or more. The adjusted Davis equation is appropriate for intermodal trains, particularly those with double-stack containers or mixtures of different intermodal car types, namely TOFC, single-stack container-on-flat-car units, and double-stack container-on-flat-car units (15). A combination of the modified and adjusted rail car resistance equation is

$$R_C = K_{\text{adj}} (0.6W_C + 20A_C + 0.01VW_C + KN_CV^2) \quad (13)$$

where

$R_C$  = adjusted rail car resistance,  
 $K_{\text{adj}}$  = adjustment factor to modernize the Davis equation,  
 $W_C$  = total weight of all cars,  
 $A_C$  = total number of axles of all cars,  
 $N_C$  = number of cars, and  
 $K$  = drag coefficient (0.07 for conventional equipment, 0.0935 for containers, and 0.1600 for trailers on flatcars).

Total train resistance ( $F_u$ ) is therefore equal to

$$F_u = 1.3W_L + 29A_L + 0.03W_LV + 0.3N_LV^2 + K_{\text{adj}} (0.6W_C + 20A_C + 0.01VW_C + KN_CV^2) \quad (14)$$

CTRail automatically varies  $K$  and  $K_{\text{adj}}$  according to the equipment selected by the user.

### Grade Resistance

Grade resistance is taken as 20 lb/ton per percent of grade. It is derived from a relationship between the angle of ascent (or descent) and gravitational forces acting on the train (14). The number 20 is a result of the conversion from tons to pounds. Grade resistance, train weight, and percentage grade can therefore be expressed as

$$F_g = 20Wg \quad (15)$$

where  $F_g$  is the grade resistance in pounds,  $W$  is the total gross weight of the train in tons, and  $g$  is the percentage gradient of the terrain.

### Curve Resistance

According to Avallone et al. the behavior of rail vehicles in curve negotiation is the subject of several ongoing AAR studies (14). Studies indicate that lubrication of the flange or gage face (or both) can significantly reduce train resistance on tangent tracks (14). However, for general estimates and nonlubricated rail with conventional trucks, curve resistance ( $F_c$ ) is calculated as

$$F_c = 0.8Wc \quad (16)$$

where  $W$  is the total gross weight of the train in tons, and  $c$  is the degree of curvature.

### Train Cruising Speed

Train cruising speed can therefore be found by using the equation of motion:

$$F_T - F_u - F_g - F_c = 0 \quad (17)$$

With the substitution into the above equation with the earlier defined  $F_T$ ,  $F_u$ ,  $F_g$ , and  $F_c$ , the equation of motion can be rewritten as

$$308 \text{ hp} - \left[ \frac{1.3W_L + 0.6K_{\text{adj}}W_C + (20g + 0.8c)W}{29A_L + 20K_{\text{adj}}A_C} \right] V - [0.03W_L + 0.01K_{\text{adj}}]V^2 - [0.3N_L + K_{\text{adj}}KN_C]V^3 \quad (18)$$

Solving Equation 18 iteratively results in the determination of the train's cruising speed ( $V$ ). Alternatively, if the train's maximum speed is specified, CTRail varies the HPTT ratio in order to calculate the required horsepower needed to power the train at the specified maximum speed.

### Fuel Consumption and Cost Module

Fuel consumption is calculated as a function of thermal efficiency, horsepower, and travel time. Thermal efficiency ( $\eta$ ) is defined as the ratio of work performed to energy consumed and varies between 25% and 30% for a rail diesel engine (13). To relate work and energy, the energy content of a gallon of fuel is assumed to be 138,700 British

thermal units (Btu), and work defined as the product of horsepower and time is converted via the formula 2,545 Btu = 1 horsepower hour (hp-h):

$$\eta = \frac{\text{work}}{\text{energy}} = \frac{2,545 \text{ gal}}{138,700 \text{ hp-h}} \quad (19)$$

Given a diesel engine with horsepower (hp), let  $n$  be equivalent to gallons of fuel consumed per hour:

$$\eta = \frac{2,545 \text{ hp}}{138,700 n} = \frac{0.0183 \text{ hp}}{n} \quad (20)$$

The above equation can then be solved as

$$n = \frac{0.0183 \text{ hp}}{\eta} \quad (21)$$

where  $n$  is the number of gallons of fuel consumed per hour by a diesel locomotive with horsepower (13). Since current technological innovations have also increased locomotive fuel efficiency, CTRail allows users to correctly specify efficiencies greater than 30%. To calculate the cost of fuel, the user specifies a price ( $p$ ) for a gallon of diesel fuel, and the fuel cost per hour ( $C_{fh}$ ) can be calculated as

$$C_{fh} = p * n \quad (22)$$

The total fuel cost per trip may be found by multiplying trip time (in hours) by fuel cost per hour.

Therefore, given trip time  $T$ , the fuel cost for a trip can be calculated as  $C_F$ , where 0.0183 hp/ $\eta$  is the relationship between gallons of fuel consumed per hour and the thermal efficiency of a diesel engine with horsepower:

$$C_F = p * \frac{0.0183 \text{ hp}}{\eta} * T \quad (23)$$

### Locomotive Emissions Module

CTRail's emissions model is based on five emission standards set by the Environmental Protection Agency (EPA). Since a single locomotive's emissions rate varies throughout its life as the engine ages and as ambient conditions change, the emission rates provided by the Environmental Protection Agency are approximations based on simplified assumptions (16). Locomotive emissions are calculated by multiplying the Environmental Protection Agency locomotive line haul emissions factors in grams per gallon by the gallons of fuel consumed. Estimates include oxides of nitrogen, particulate matter, and carbon monoxide.

### Crew Labor Cost Module

Formulas based on distance traveled have been used to calculate crew wages. This approach, though sometimes appropriate, may not necessarily be accurate as different railroads have different formulas for determining crew wages. To simplify this process, the model currently enables users to specify a fixed rate per 8-h shift for each crew member. This number is then multiplied by the number of crew

members. Future enhancements of the model will give users more flexibility in calculating line haul labor costs.

### Maintenance Cost Module

Track maintenance cost is determined by multiplying the per mile system average rate ( $c_{mr}$ ) by the number of cars and locomotives in operation. Car maintenance cost is specified by the user on a per mile ( $c_{mc}$ ) basis multiplied by the number of cars in operation. Locomotive maintenance cost is also specified by the user on a per mile value ( $c_{ml}$ ) basis and multiplied by the number of locomotives in operation. Total maintenance cost ( $C_M$ ) is calculated as

$$C_M = c_{mr}(N_C + N_L) + c_{mc}N_C + c_{ml}N_L \quad (24)$$

Current CTRail estimates are based on rail expert recommendations and may not be necessarily accurate for each individual railroad. However, with the integration of the asset management module, public stakeholders will be able to develop more accurate maintenance figures based on the railroad's anticipated maintenance expenditures. These can be calculated as a function of locomotive miles and car miles moved annually as well as the cost associated with maintaining the rail tracks (17–20).

### Capital Cost and Investment Cost Module

Capital and investment costs are the most difficult to model. Railway capital costs include large investments in the construction of rail tracks, structures, railyards, signals, and car and locomotive purchases. Without sufficient and reliable data, modeling the investment cost associated with rail tracks, structures, railyards, and signals is almost impossible. CTRail therefore only accounts for investment costs associated with locomotive and car purchases. These are known as the locomotive ownership cost and the car ownership cost. Using a straight-line depreciation equation, depreciation charge per hour is determined and multiplied by the total trip time as shown in Equation 25:

$$\begin{aligned} \text{hourly depreciation} &= \frac{\text{cost of asset} - \text{scrap value}}{\text{life span (years)} \times 8,760 \frac{\text{h}}{\text{years}}} \\ &\times \text{trip time (h)} \times N \end{aligned} \quad (25)$$

### Model Limitations and Future Enhancements

CTRail is limited to line haul movement operation and therefore does not account for terminal operations. Terminal operations are a substantial part of railroad operations that cannot be ignored in railroad cost analysis. However, for the purposes of this research, terminal operations and costs are assumed to be the same for all origins and destinations, and loading and unloading operational costs are included to account for economies of scale in line haul operation.

Capital investments such as road construction, right-of-way acquisition, grading, signal and interlock installation, stations and office buildings, and all other infrastructural investment costs are not included because of lack of sufficient supporting data and variability among the various rail companies. Other expenses ignored

include equipment rentals, purchased services, and other indirect expenses.

Other operational limitations include an assumption of average speed. Variable speed caused by changes in grade, curvature, wind resistance, and traffic delays can be modeled with CTRail, but this capability is not discussed in this paper. According to rail experts, trains are operated at full throttle whenever possible, and acceleration and deceleration calculations are also omitted because of their relative insignificance in comparison to the entire trip. For fuel consumption, CTRail assumes the train is running at full throttle. The model also assumes all identical locomotives run at the same horsepower. This might not necessarily be the case as railroad companies run locomotives at different horsepower to optimize fuel consumption or enhance tractive effort. Future enhancement of the model would enable users to choose multiple locomotives of varying horsepower.

Finally, there are insufficient data from rail companies to enable modelers to adequately estimate the capital, maintenance, and administrative costs associated with each trip, thereby making the determination of actual prices almost impossible. Railroads are reluctant to share such data because of the competitive nature of the business. Depending on the commodity type, market dominance, and the route being used, railroad companies have additional charges such as switch charges, hazmat, and other charges not currently captured in the model. In addition, railroads install and maintain traffic signals, construct sidings, develop double tracks, and spend on other capital investments that cannot be captured by this model. Because of these various limitations, it is advised that CTRail be used for rail cost comparison purposes only and not for determining railroad rates. CTRail is currently not a complete rail analysis model and would need to be integrated with the other components of the rail analysis modeling framework.

### CASE STUDIES

Two case studies are performed to demonstrate the capabilities of CTRail. The case studies involve the transport of goods between Houston, Texas, and Dallas–Fort Worth (DFW), Texas, and examine various scenarios in an effort to identify conditions that would make rail transport competitive with trucking for short-haul distances. Trade between Houston and DFW is projected to continue increasing in the years ahead (21). Houston and DFW, together with Austin and San Antonio, Texas, form part of the Texas triangle, which accounts for the majority of commodities originating from or destined for Texas (22).

The first study attempts to determine a threshold at which a certain quantity of goods when transported justifies the need for an efficient short-haul rail infrastructure. This study is necessary to assist policy makers in deciding whether it is worthwhile to invest in short-haul freight rail based on the quantity of goods transported between the origin and destination points. The second study examines if faster speeds can make freight rail more competitive with trucking and still be cost-effective.

#### Case Study 1. Quantity of Goods and Short-Haul Freight Movement

According to 2002 Freight Analysis Framework data, 326,000 tons of exported goods were transported via truck or rail from Dallas to Houston. This number was expected to increase by 100% in 2010, 200% by 2015, and more than 500% by 2025. Should these projections be

accurate, 978,500 tons of goods will be transported in 2015 and more than 2 million tons of goods by 2025. With the 2002 Freight Analysis Framework data, a study was performed to determine if it is viable for policy makers to support a more efficient freight rail system than the one currently in existence. Assuming a generous 30% adoption rate for rail movements, the number of containers for each projected year was calculated, assuming each container carries 15 tons of cargo.

Rail inputs based on rail expert assumptions were used for the freight train. The number of containers was varied from 50 to 200 at 10-container intervals to determine the optimum number of containers that could be transported on a single trip (although some trains operated at 300 or more containers in 2009 and 2010). Rail inputs are as follows:

- Track maintenance: \$0.50 per car per locomotive mile,
- Loading and unloading cost: \$150.00 per container,
- Car maintenance: \$0.12/mi,
- Cargo weight per container: 15 tons,
- Locomotive maintenance: \$2.10/mi,
- Fuel price: \$2.00/gal,
- Average crew wages: \$360 per 8-h shift per crew member,
- Maximum train speed: 55 mph,
- Number of crew members: two,
- Utilization ratio: 100%,
- Tare weight of 40-ft container: 4.2 tons,
- Distance: 275 mi,
- Tare weight of single well: 17.60 tons,
- Configuration: double stacked, and
- Locomotive horsepower: 4,400.

The result of this analysis is illustrated in Figure 1, which shows that the optimum number of containers that can be transported by rail can be either 110 containers at 8.10 cents/ton-mile (Train 1) or 220 containers at 8.03 cents/ton-mile (Train 2). A sample breakdown of operating costs also showed that 70% of total cost was loading and unloading cost (terminal operations), 22% maintenance, 7% fuel, 1% crew wages, and 0.2% capital costs.

Using the two parameters obtained from Figure 1, the number of annual train trips needed to meet projected demand was then calculated. The number of truck trips was also calculated for two scenarios: transporting 40-ft containers and transporting 53-ft containers. CTRail calculated that 1,568 gal of fuel is consumed on a single trip

by a train hauling 110 containers (Train 1) and 3,136 gal of fuel by a train hauling 220 containers (Train 2) (an additional locomotive is needed in this case). Trucks are assumed to consume fuel at an average rate of 5.5 mpg.

Figure 2 illustrates that as the number of containers increases over the years, the amount of fuel consumed by the trucks increases at a much faster rate than that of the two trains. Fuel consumption increased from less than 500,000 gal in 2002 to more than 2.5 million gallons in 2035 for trucks hauling 53-ft containers, and more than 3.5 million gallons for trucks hauling 40-ft containers. Even though Train 1 does twice as many trips as Train 2, it uses 50% less fuel on each trip, so both trains end up consuming the same amount of fuel for the same quantity of goods transported, as shown by the overlapping lines in Figure 2. The fuel consumption for each of these trains is less than 1.2 million gallons in 2035. The cost associated with transporting the containers (not shown here) increased in a fashion similar to that shown in Figure 2.

The results of Case Study 1 suggest that for large quantities of commodities, the advantages of line haul rail outweigh those of trucking (Figure 2). Rail is more cost-effective and fuel efficient than trucking in the long term, thus meriting rail investment by policy makers.

## Case Study 2. Varying Delivery Times

In this case study, the sensitivity of faster speeds and associated costs were modeled to determine if rail can move at faster speeds and still have lower costs, making it more competitive with trucking. Speed influences fuel costs, track maintenance costs, crew costs, and hourly capital and investment costs. Inputs similar to those of Case Study 1 were used, and the numbers of containers moved varied from 50 to 200 containers. The baseline ton-mile cost for direct highway trucking based on industry estimates for the specified distance was calculated to be 7.7 cents/ton-mile.

### Trailer-on-Flat-Car Scheme

In this scenario a TOFC scheme was simulated. Figure 3 shows that even at high speeds of 80 mph, freight rail can still be cheaper than trucking on a cost per ton-mile basis when there is little or no

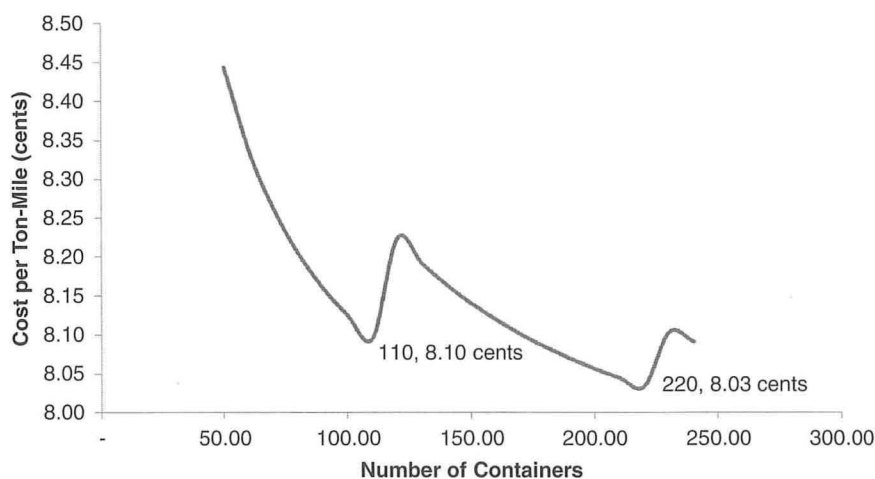


FIGURE 1 Optimum number of containers that can be transported on single trip.

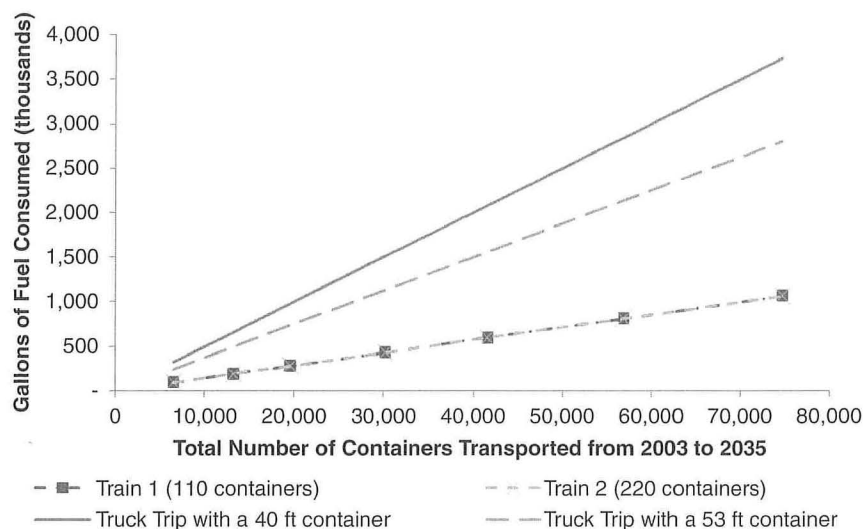


FIGURE 2 Case Study 1: gallons of fuel consumed.

difference between loading and unloading costs per container between truck and rail. However, when high terminal costs for loading and unloading of the intermodal rail car are added, the line haul cost advantage over trucking diminishes for TOFC trains (Figure 4).

#### Double-Stacked Intermodal Scheme

In this scenario a double-stacked intermodal scheme was simulated to examine the advantages of double stacking. For double-stacked intermodal trains, freight rail remains cheaper than trucking even when the loading and unloading cost difference between truck and rail is as high as \$200 per container (Figure 5). This reinforces the observation made by Resor et al. that for intermodal rail to be competitive for short-haul distances, it is critical that terminal operations and drayage costs be very low (23). Low drayage cost will enable intermodal rail to be competitive with trucking at short-haul distances and even at speeds greater than 65 mph, the average speed of trucks on rural freeways.

#### CONCLUSIONS

CTRail seeks to replace the less-flexible uniform rail costing model and also provides a tool for policy makers and freight transport stakeholders to make their own independent freight movement analysis involving railways. In part, CTRail is based on a simulation of the horsepower required to move a certain ton of cargo at a specified speed over a certain distance. The calculated horsepower is used to determine the fuel consumed, which is translated into fuel cost and emissions generated. Maintenance costs are assumed to be variable costs calculated according to the number of equipment units (locomotives and rail cars) used and the distance traveled. Crew wages are calculated as a daily rate determined by the travel time and capital cost and shown as an hourly input rate. The model also allows multiple user-specified inputs.

Despite its flexible and route- and corridor-specific analytical capabilities, the current version of CTRail limits the types of analysis that can be performed. These limitations will be addressed in

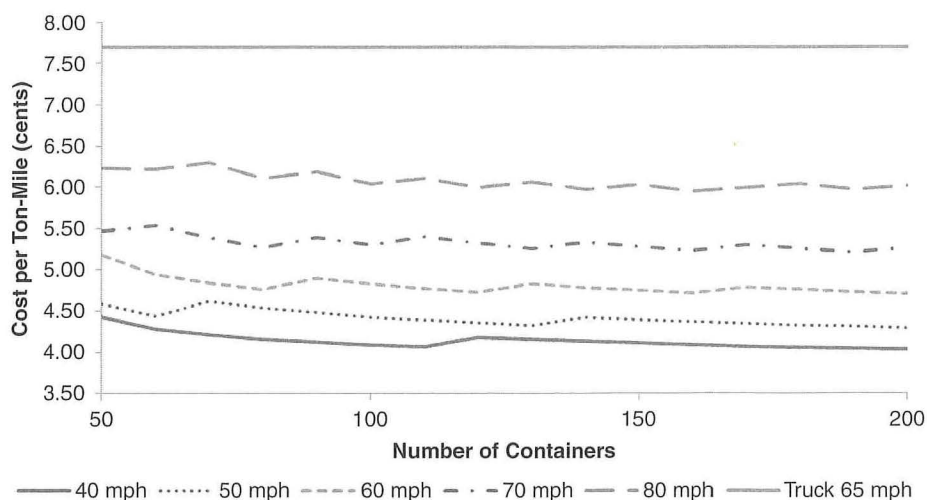


FIGURE 3 Ton-mile costs for TOFCs with no difference in loading and unloading costs per container between truck and rail and rail travels at different speeds.

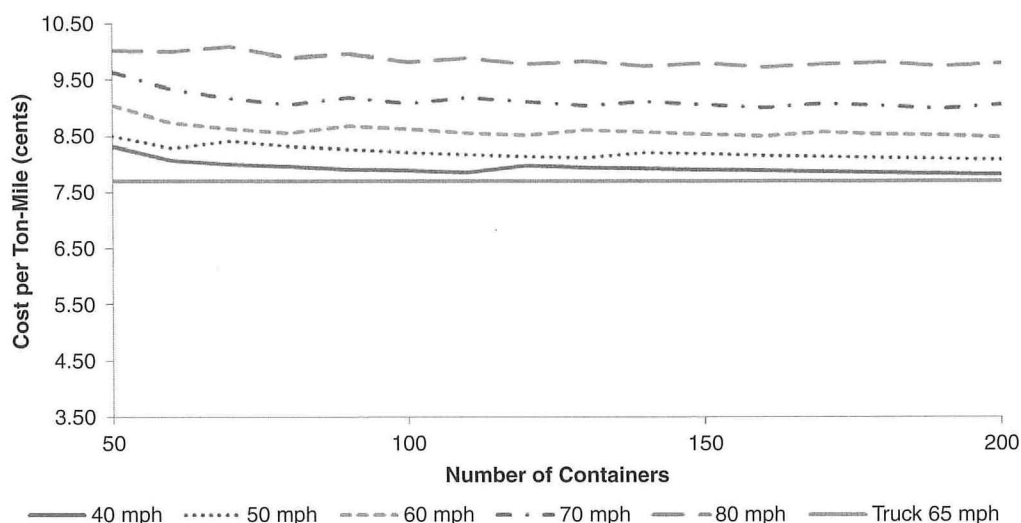


FIGURE 4 Ton-mile costs for TOFCs with \$200 difference in terminal charges (\$100 loading and \$100 unloading) per container between truck and rail and rail travels at different speeds.

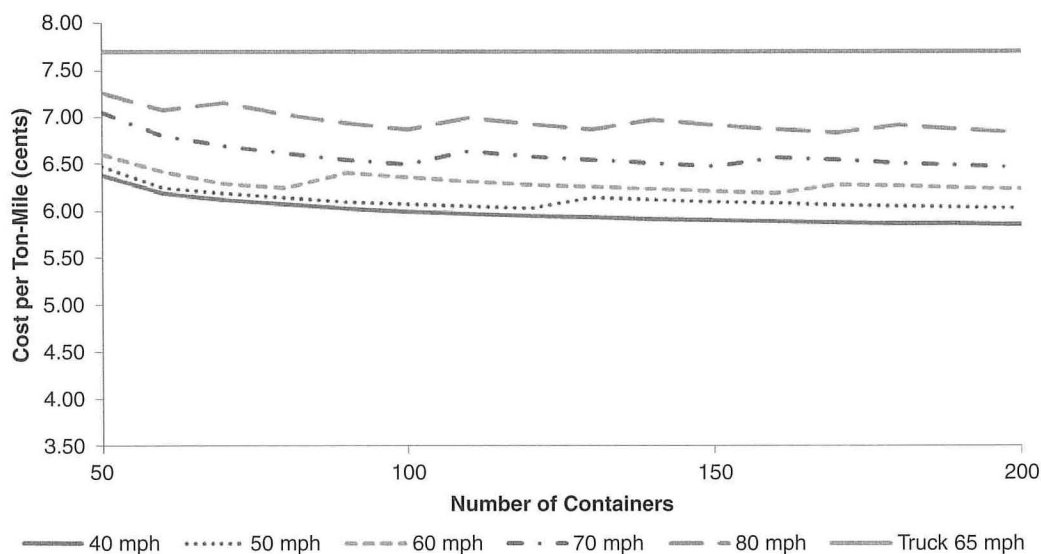


FIGURE 5 Ton-mile costs for double-stacked containers with \$200 difference in terminal charges (\$100 loading and \$100 unloading) per container between truck and rail and rail travels at different speeds.

future versions of the model. Upcoming versions will seek to enhance the simulation of train movements and provide the ability to capture the effects of traffic volume on unit cost. The model could also be developed to include many social costs currently ignored in transportation evaluations, such as noise, air quality, accidents, extended pavement life, and carbon footprints. There is little doubt that when these are included in modal choice, rail becomes more attractive in general, and the breakeven distance at which rail becomes competitive with trucks decreases. An evaluation of the model in more detail with a Class I railroad is planned.

#### ACKNOWLEDGMENTS

The authors acknowledge constructive comments and encouragement from Carl Martland of the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, and William

Huneke and Michael Smith of the Surface Transportation Board on an earlier version of the model.

#### REFERENCES

1. Forkenbrock, D. J. Comparison of External Costs of Rail and Truck Freight Transportation. *Transportation Research Part A: Policy and Practice*, Vol. 35, No. 4, 2001, pp. 321–337.
2. Bereskin, C. G. Railroad Economies of Scale, Scope and Density Revisited. *Journal of the Transportation Research Forum*, Vol. 48, No. 2, 2009, pp. 23–38.
3. Bereskin, C. G. Sequential Estimation of Railroad Costs for Specific Traffic. *Transportation Journal*, Vol. 40, No. 3, 2001, pp. 33–45.
4. Bereskin, C. G. Railroad Capital Stock Changes in the Post-Deregulation Period. *Journal of the Transportation Research Forum*, Vol. 46, No. 1, 2007, pp. 95–110.
5. Caves, D. W., L. R. Christensen, and J. A. Swanson. The High Cost of Regulating U.S. Railroads. *Regulation*, Vol. 5, No. 1, 1981, pp. 41–46.

6. Caves, D. W., L. R. Christensen, and J. A. Swanson. Productivity in U.S. Railroads, 1951–1974. *Bell Journal of Economics*, Vol. 11, No. 1, 1980, pp. 166–181.
7. Ivaldi, M., and G. J. MacCullough. Density and Integration Effects on Class I U.S. Freight Railroads. *Journal of Regulatory Economics*, Vol. 19, No. 2, 2001, pp. 161–182.
8. Spady, R. H., and A. F. Friedlaender. *Economic Estimation of Cost Functions in the Transportation Industries*. Report No. 76–13, Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge, 1976.
9. Uniform Rail Costing System. <http://www.stb.dot.gov/stb/industry/urcs.html>. Accessed April 25, 2009.
10. AECOM Consult Incorporated. *Cost-Allocation Methods for Commuter, Intercity, and Freight Rail Operations on Shared-Use Rail Systems and Corridors*. NCHRP Project 20-65. Transportation Research Board, Washington, D.C., 2007.
11. Smith, M. E. Demonstration of the Rail Energy Cost Analysis Package: The Route Perspective (RECAP II). In *Transportation Research Record 1006*, TRB, National Research Council, Washington, D.C., 1985, pp. 23–31.
12. Hay, W. W. *Railroad Engineering*. John Wiley & Sons, New York, 1982.
13. DeSalvo, J. S. A Process Function for Rail Line-Haul Operations. Rand Corporation, Santa Monica, Calif., 1967.
14. Avallone, E. A., T. Baumeister, and A. Sadegh. *Marks' Standard Handbook for Mechanical Engineers*, 11th ed. McGraw-Hill, New York, 2006.
15. RailSim. <http://www.railsim.com>. Accessed June 25, 2009.
16. Office of Transportation and Air Quality, Environmental Protection Agency. *Emission Factors for Locomotives*. EPA-420-F-09-02. <http://www.epa.gov/OMS/regs/nonroad/locomotv/420f09025.pdf>. Accessed July 18, 2009.
17. Higgins, A. Scheduling of Railway Track Maintenance Activities and Crews. *Journal of the Operational Research Society*, Vol. 49, No. 10, 1998, pp. 1026–1033.
18. Johansson, P., and J.-E. Nilsson. An Economic Analysis of Track Maintenance Costs. *Transport Policy*, Vol. 11, No. 3, 2004, pp. 277–286.
19. Ferreira, L., and M. H. Murray. Modeling Rail Track Deterioration and Maintenance: Current Practices and Future. *Transport Reviews*, Vol. 17, No. 3, 1997, pp. 207–221.
20. Dekker, R. Applications of Maintenance Optimization Models: A Review and Analysis. *Reliability Engineering and System Safety*, Vol. 51, No. 3, 1996, pp. 229–240.
21. FHWA. *Freight Analysis Framework*. U.S. Department of Transportation, 2002.
22. Prozzi, J., K. Spurgeon, and R. Harrison. Secret Lives of Containers: Evidence from Texas. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1833, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 3–10.
23. Resor, R. R., J. R. Blaze, and E. K. Morlok. Short-Haul Rail Intermodal: Can It Compete with Trucks? In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1873, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 45–52.

---

*The Local and Regional Rail Freight Transport Committee peer-reviewed this paper.*

# Noncontact Ultrasonic Guided-Wave System for Rail Inspection

## Update on Project at University of California, San Diego

Stefano Coccia, Robert Phillips, Ivan Bartoli, Salvatore Salamone, Francesco Lanza di Scalea, Mahmood Fateh, and Gary Carr

The University of California, San Diego (UCSD), with an FRA Office of Research and Development grant, is developing a system for high-speed and noncontact rail defect detection. A prototype was designed and field tested with the support of Volpe National Transportation Systems Center and ENSCO, Inc. The goal of this project was to develop a rail defect detection system that provided (a) better defect detection reliability (including internal transverse head defects under shelling and vertical split heads) and (b) higher inspection speed than achievable by current rail inspection systems. This effort was also in direct response to safety recommendations issued by the National Transportation Safety Board after the disastrous train derailments at Superior, Wisconsin, in 1992 and Oneida, New York, in 2007, among others. The UCSD prototype used noncontact ultrasonic probing of the rail head (laser and air-coupled sensors), ultrasonic guided waves, and a proprietary real-time statistical analysis algorithm that maximized the sensitivity to defects while it minimized false positives. The design allowed potential inspection speeds up to 40 mph, although to date all field tests were conducted up to 15 mph. This paper (a) summarizes the latest technology development test conducted at the rail defect farm of Herzog, Inc., in Saint Joseph, Missouri, in June 2010 and (b) describes the completion of the new rail defect farm facility at the UCSD Camp Elliott Field Station with partial in-kind donations from the Burlington Northern Santa Fe Railway.

Conventional ultrasonic rail inspection uses piezoelectric transducers that are coupled to the top of the rail with ultrasonic wheels or sleds filled with water or other fluids (1). The most serious drawback of this method is that shallow surface cracks (shelling) can mask internal transverse defects (TDs). This limitation was the cause of train derailments in Superior, Wisconsin, in 1992 and Oneida, New York, in 2007, where severe problems were caused by hazardous material spillage. In response to these and other accidents, the National Transportation Safety Board issued safety recommendations to FRA for

improving the effectiveness of rail inspection technologies to detect internal rail defects, particularly under shelling (2, 3). Other drawbacks of wheel-based ultrasonic rail inspections are limited speed (typically less than 15 mph) and challenges in detecting vertical split head (VSH) defects, also critical for rail safety.

FRA safety statistics data show that train accidents caused by track failures, including rail, joint bar, and anchoring failures, resulted in 3,386 derailments and \$685 million in associated damage costs in the decade from 1998 to 2008 (4). The leading cause of these accidents was the transverse fissure defect, which was responsible for 815 derailments and \$160 million in costs during the same decade (Figure 1). Another type of TD, the rolling contact fatigue defect, typically initiates at the gage corner of the railhead. The detail fracture, the most common rolling contact fatigue defect, was responsible for 427 derailments and \$137 million in associated damages (second-highest cost) from 1998 to 2008 in the United States (Figure 1).

In response to these statistics, the primary targets of the University of California, San Diego (UCSD)–FRA rail inspection prototype are TDs (transverse fissures and detail fractures), including those under shelling, as well as VSHs and compound fractures.

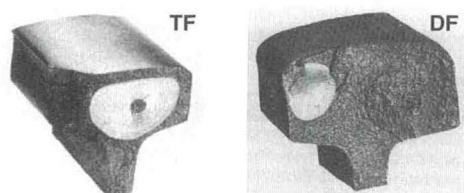
The UCSD–FRA system uses noncontact transduction of ultrasonic waves in the rail head (laser and air-coupled sensors) (5). Lift-off distances for the sensors are on the order of 2 in. from the top of the rail head. The system also uses a proprietary signal-processing algorithm based on statistical analysis that maximizes the defect indications and minimizes false positive indications. Unlike other rail ultrasonic systems that use bulk waves (longitudinal or shear), the UCSD–FRA system uses guided ultrasonic waves. The ultrasonic guided modes insonify a large portion of the railhead and allow for a longer gage length which, in turn, increases the achievable inspection speed. The prototype has been tested at speeds up to 15 mph in the field, although higher speeds are potentially possible. The maximum speed potentially achievable with the current design is on the order of 40 mph. Higher speeds would require some modifications to the hardware design. Figure 2 shows pictures of the prototype towed by the FRA R-4 Hy-railer during a field test.

### RESULTS OF BLIND FIELD TESTS AT HERZOG

This section summarizes the results of two blind tests conducted during the June 2010 technology development tests at Herzog Services, Inc., in Saint Joseph, Missouri. ENSCO, Inc. provided field test

S. Coccia, R. Phillips, I. Bartoli, S. Salamone, and F. Lanza di Scalea, Nondestructive Evaluation and Structural Health Monitoring Laboratory, University of California, San Diego, 9500 Gilman Drive, M.C. 0085, La Jolla, CA 92093-0085. M. Fateh and G. Carr, Office of Research and Development, Federal Railroad Administration, 1200 New Jersey Avenue, SE, Washington, DC 20590. Corresponding author: F. Lanza di Scalea, flanza@ucsd.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 143–147.  
DOI: 10.3141/2261-16



Type of Defect	% Total Defects	Direct Damage Cost	# Derailments
Transverse/Compound Fissure	23 % (1 <sup>st</sup> leading cause)	\$ 160 M (highest cost)	815
Detail Fracture	12 % (2 <sup>nd</sup> leading cause)	\$ 137 M (2 <sup>nd</sup> highest cost)	427

FIGURE 1 Transverse fissure (TF), detail fracture (DF), and FRA safety statistics data for rail, joint bar, and rail anchoring on all U.S. railroads, 1998 to 2008 (M = million).

support. Figure 3 shows pictures of the prototype and some of the test participants.

The test track included 12 railhead defects, including detail fractures, TDs under shelling, defective field and plant welds, side-drilled holes (simulating TDs), and horizontal and vertical split head defects. Both blind tests were conducted at low speed (~2 mph) and mostly on tangent 136 rail extension track.

Two signal-processing approaches were used for the two tests. One configuration was less sensitive to small railhead discontinuities (Blind Test 1) than the other (Blind Test 2).

Ten of 12 defects were correctly detected by Blind Test 1 (less sensitive configuration), and 11 of 12 defects were correctly detected by Blind Test 2 (more sensitive configuration). Blind Test 1 therefore had an 83.34% detection rate with zero false positives. Blind Test 2 had a 91.67% detection rate at the cost of four false positives. However, after hand-mapping of the test area, three of the four false positives mapped to shallow defects under shelling. This made the true false positive detection for Blind Test 2 as low as one false positive.

The UCSD list of defects provided for Blind Tests 1 and 2 did not include an 80% defective plant weld. Although this defect was correctly detected by the system, it was not included in the original list because it is considered a weld. Later in the tests a method was identified to distinguish good welds from defective welds according to their different ultrasonic signatures. When this differentiation was applied, the 80% defective plant weld was consistently detected as a defect by the system.

Figure 4 plots the results of the two blind tests along with the industry average and American Railway Engineering and Maintenance-of-Way Association (AREMA) recommendations for reliability of defect detection (particularly TDs). Detection percentages were computed as the number of detected defects divided by the total number of defects of a given size class. The 21% to 40% size class was not tested since no defect of such size was present on the track. The plot shows that the UCSD system outperformed both industry average and AREMA recommendations in all defect size classes, including the largest size class of 81% to 100%, once the weld differentiation method was implemented.

### Influence of Rail Surface Conditions

The Herzog test track contained different levels of railhead surface discontinuities, including shells and head checks. The UCSD system identifies defects as peaks in a damage index plot that is computed and plotted in real-time at each position along the rail by the statistical signal-processing algorithm. Different scales of visualization of the damage index plots were used for different areas of the rail. This was done to adapt the response of the system to different rail surface conditions. The defect-free level of the damage index varied in response to the different rail surface conditions encountered in the tests, but the defect indications were still distinguishable from the noise level. An automatic gain control feature, similar to that used in common ultrasonic rail inspections, should be implemented in the final configuration of the system to rescale the data in the presence, for example, of moderate and heavy shelling.

However, the system's sensitivity to different rail surface conditions could also be potentially useful in estimating the severity of



FIGURE 2 UCSD-FRA rail inspection prototype towed by FRA R-4 Hy-railer.



FIGURE 3 UCSD prototype at Herzog (*top*) and some test participants from UCSD, ENSCO, FRA, Volpe, Herzog, BNSF, and Union Pacific (*bottom*).

shelling, and hence the thickness of the layer that requires grinding during rail maintenance. Figure 5 shows one of the test runs at Herzog over a section of rail with head checks. Although the damage index plot is sensitive to the surface condition of the rail, the 10% head area defective field weld at position 87 ft 2 in. is still well recognizable over the noise floor.

### Weld Signatures

One achievement of the study was the recognition that the system responded differently to good welds and defective welds, hence allowing for the detection of weld defects. This differentiation is not always achievable by current ultrasonic rail inspection systems because the coarse-grained structure of welds often prevents the penetration of high-frequency ultrasonic beams. The prototype detected a good weld at 54 ft 6 in., a 10% TD at 55 ft 0 in., and a 50% defective field weld at 59 ft 2 in. (Figure 6). As the plot shows, the response to a good weld is a high-level stable plateau with no local minima points, while defects and defective welds produce a jumper damage index with several local minima points. This behavior could be used to train an automatic defect classification algorithm or used under operator's judgment to detect defects within welds.

### Detection of Vertical Split Head

The prototype's primary goal is the detection of TDs, which are historically the number one cause of concern for train accidents related to rail defects. The number two cause of concern is VSH. The tests at Herzog demonstrated that the system's reliability in detecting VSHs is excellent. This is an important achievement, since VSHs are often missed by conventional ultrasonic rail inspections because their orientation may not generate a strong reflection of the ultrasonic beam from ultrasonic wheel search units. Figure 7 shows an example of detection of a 1-ft-long VSH present at Herzog's rail defect farm.

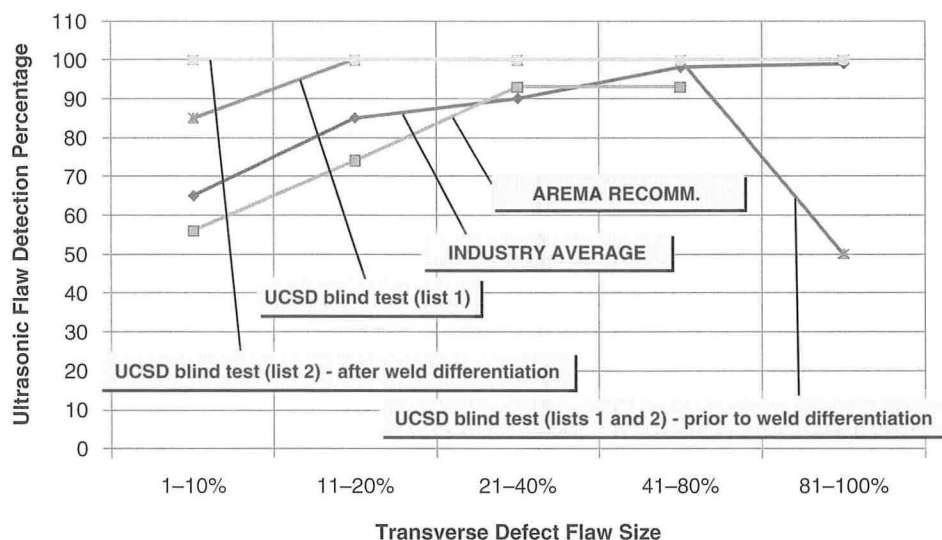


FIGURE 4 Results of UCSD blind tests at Herzog rail defect farm on June 15, 2010, compared with industry average and AREMA standards (recomm. = recommendation).

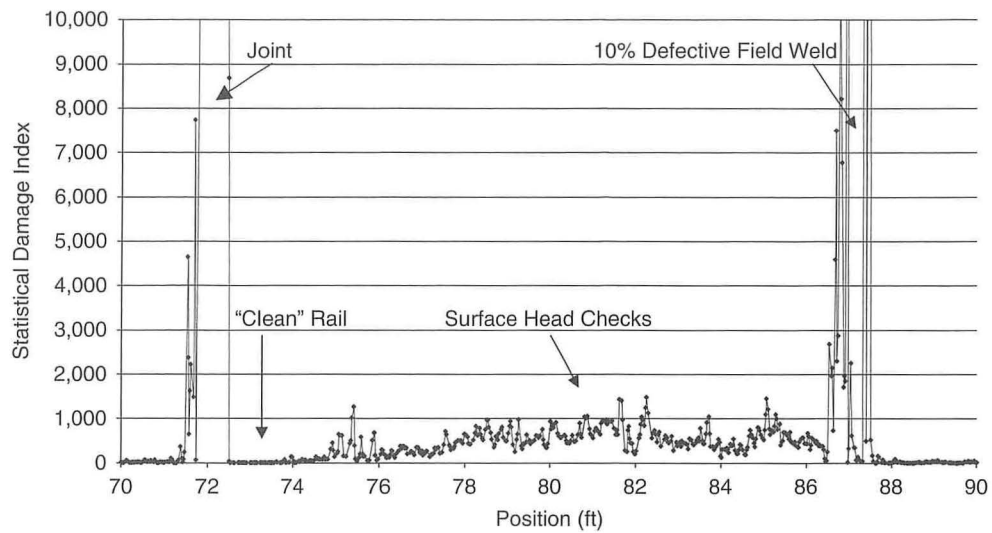


FIGURE 5 Damage index plot of test run conducted at 2 mph.

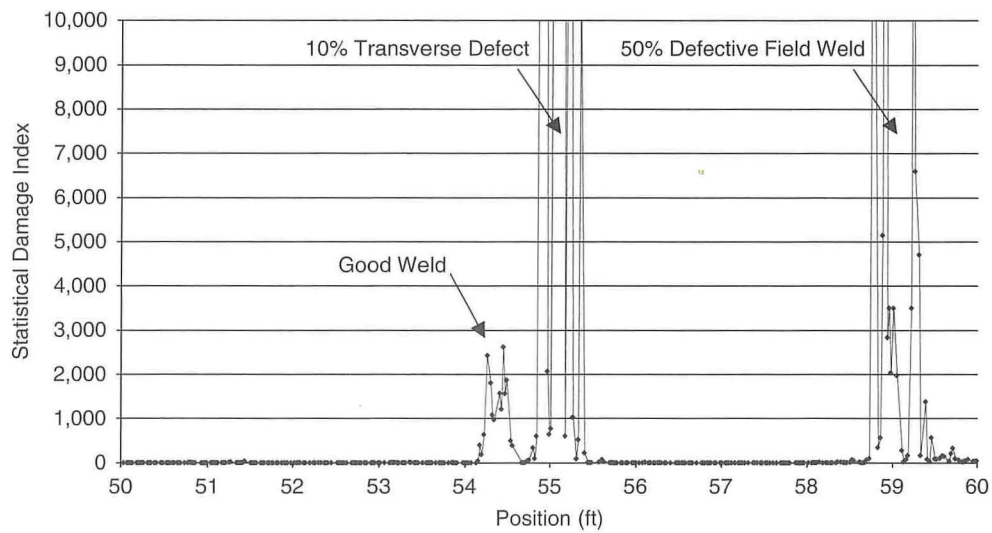


FIGURE 6 Examples of different signatures of good weld, TD, and defective weld.

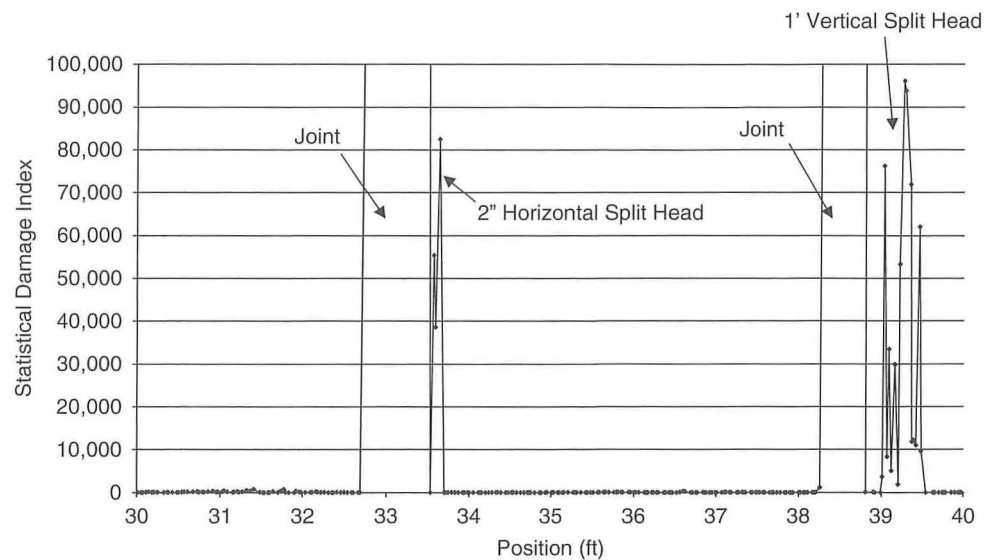


FIGURE 7 Example of clear detection of 1-ft-long VSH (two joints and horizontal split head also shown as detected).



FIGURE 8 New rail defect farm at UCSD for development of rail inspection technologies.

### NEW UCSD RAIL DEFECT FARM

It was determined in 2009 that further development of the rail inspection prototype required a new defect farm on site at UCSD. This facility has now been constructed at the UCSD Camp Elliott Field Station, about 8 mi from the main UCSD campus. The Camp Elliott Field Station also hosts some of UCSD's high-visibility structural testing facilities, including the world's only outdoor shake table for earthquake engineering testing and a blast simulator facility for blast studies.

The new UCSD–FRA rail defect farm is a 250-ft-long track with a tangent portion and an 8° curved portion (Figure 8). Burlington Northern Santa Fe (BNSF) donated the tracks and most of the ties and ballast. Sopac Rail, Inc. performed the construction. The track features about 15 natural rail defects, including TDs under shelling and some artificial rail defects. This facility will be used primarily for the technology development of the UCSD–FRA rail inspection prototype. It is also expected that the facility will be available to other developers of rail inspection technologies.

### DISCUSSION AND CONCLUSIONS

The performance of the UCSD–FRA rail inspection system at Herzog's rail defect farm was very promising. Two blind tests were conducted at slow speed (~2 mph). The system reliably detected TDs, including some under shelling, side-drilled holes, artificial vertical and horizontal split heads, and defective field and plant welds. The defect detection reliability shown during the blind tests exceeded industry average and AREMA recommendations.

The system was also sensitive to the presence of good welds, which showed a different signature than the one related to the flaws.

Testing at higher speed (up to the 9 mph allowed on the Herzog test track) was also conducted after the blind tests. The system performed well at these speeds, although with a decreased position resolution compared to the lower speeds. Modifications to the system hardware should be made to achieve robust performance at the higher speeds.

An interesting outcome of the Herzog tests was the excellent detectability of VSH and the consequent potential for characterizing different rail surface conditions, which could be useful for improving the scheduling of rail grinding maintenance. However, the VSH at the Herzog defect farm was an artificial, man-made defect. VSH or rail shear defects developing from rail manufacturing processes or caused by fatigue may produce different results. The signal-to-noise ratio of the defect indications was very satisfactory. Clearly, a more robust assessment of the defect detection reliability of the system will require testing on a larger variety of defects.

The recently completed UCSD rail defect farm facility, a 250-ft-long track with a number of artificial and real defects built with FRA funding and BNSF in-kind support, will be available for technology development of the UCSD rail inspection system as well as available to other developers of rail inspection technologies.

### ACKNOWLEDGMENTS

This work was supported by grants from the U.S. Federal Railroad Administration. The National Science Foundation funded the initial research effort. John Choros of the Volpe Center participated in the field tests at Herzog as adviser and evaluator. ENSCO, Inc., provided field support for these tests. Special thanks are extended to Troy Elbert of Herzog, Inc., for providing access to the rail defect farm and to Scott Staples of BNSF for arranging for the donation of materials for the rail defect farm in San Diego.

### REFERENCES

1. Lanza di Scalea, F. Ultrasonic Testing Applications in the Railroad Industry. In *Nondestructive Testing Handbook* (P. O. Moore, ed.), 3rd ed. American Society for Nondestructive Testing, Columbus, Ohio, 2007, pp. 535–552.
2. *Hazardous Materials Release from Railroad Tank Car with Subsequent Fire at Riverview, Michigan, July 14, 2001*. Report HZM-94/01. National Transportation Safety Board, 2001.
3. *Derailed of CSX Transportation Train No. Q39010, Oneida, New York, March 12, 2007*. Report RAB-08/05. National Transportation Safety Board, 2007.
4. FRA, U.S. Department of Transportation. *Safety Statistics Data: 1998–2008*. <http://safetydata.fra.dot.gov/OfficeofSafety/default.aspx>.
5. Coccia, S., I. Bartoli, R. Phillips, S. Salamone, F. Lanza di Scalea, M. Fateh, and G. Carr. UCSD/FRA Ultrasonic Guided-Wave System for Rail Inspection. *Proc., AREMA Annual Conference*, Chicago, Ill., 2009.

*The Railway Maintenance Committee peer-reviewed this paper.*

# Installation of Railroad Wayside Defect Detectors

## Multiperiod Design

Fan Peng, Xiaopeng Li, and Yanfeng Ouyang

Railroads use wayside inspection technologies to monitor the health of passing railcars. Because of resource constraints, an efficient and feasible installation plan should be integrated into a multiperiod decision-support framework to derive the maximum benefit over a long planning horizon. A large-scale network optimization framework was proposed to solve this problem. The problem was formulated into two equivalent mathematical models: a maximum coverage model and a  $K$ -median model. A set of solution algorithms including a greedy heuristic, a 1-interchange heuristic, and a Lagrangian relaxation method were developed for these models, and their computational performances were investigated. A numerical case study was developed to illustrate the use of these models in real-world settings, and insights into empirical applications of these models were drawn.

Sensor technologies have been widely applied in railroad transportation networks. A variety of technologies, generally known as wayside detectors, are installed adjacent to railroad tracks to monitor the health and performance of passing railcars (1). Wayside detectors were developed decades ago; examples include the dragging equipment detectors developed in the 1930s (2–4) and the hot bearing detectors developed in the 1940s and 1950s (5, 6). These detectors are normally reactive, in the sense that they are only capable of detecting defects that have a short latency period before failure (7). When an alert from a reactive system is received, regular train operation has to be disrupted in order for maintenance crews to take actions immediately, or very soon, to prevent possibly catastrophic consequences. Reactive detectors need to be densely deployed (e.g., spaced approximately every 20 mi) over the railroad network to ensure that all railcars can be frequently inspected.

More recent detectors rely on advanced technologies and have dramatically different working mechanisms. They are sometimes referred to as predictive or preventive detectors (7). Examples include rail bearing acoustic monitors (8) and truck performance detectors (9, 10). Preventive systems predict future defects or failures on the basis of detailed performance records, so that advance notice can be issued ahead of any imminent failures. Railcars only need to be

inspected by predictive detectors once in a long while (e.g., 6 months or more), and preventive maintenance can be scheduled at times when railcars are not in regular operation. Hence, predictive technologies have very significant economic and efficiency advantages. However, predictive systems in general are much more expensive than reactive systems. Each predictive system installation may cost as much as three-quarters of a million dollars.

Consequently, detector deployment strategies for predictive systems differ from those for reactive systems. As a result of limited resources, railroad practitioners prefer to install predictive detectors at selected optimal locations to cover a large portion of traffic volume. However, this is not a trivial problem. For example, it was shown that simply installing detectors at the busiest locations (i.e., with the highest volume of passing traffic) tends to yield suboptimal inspection results—a small portion of railcars may pass these locations (and hence be inspected) multiple times, while traffic over the rest of the network could be missed. Hence, optimal design of predictive detector locations should consider the interdependence of installation locations that results from shared traffic flow, minimize inspection redundancy, and maximize the inspections of distinct railcars. Given the complex and entangling traffic flows in railway networks, finding such optimal installation locations imposes a huge challenge.

Generally, there is a lack of research on railroad wayside detector location design. Only recently, a systematic optimization framework for this problem was proposed by Ouyang et al. (11). Numerous studies on sensor location problems have been conducted in the highway context (12), including robust analysis on sensor utilities (13) and optimal sensor location design (14) and its various extensions (15–17). Probabilistic sensor failures and generalized surveillance benefits have been addressed by Li and Ouyang (18, 19).

Most studies assume that wayside sensors are implemented all at the same time (11). In railroad practice, however, a sensor system is more likely to be installed incrementally over a long span of time. There are several reasons for such incremental development. First, the budget available for wayside detector installations is often limited. Because of expensive installation costs (especially for advanced technologies), installation usually has to be carried out period by period. Second, traffic patterns may change dramatically from time to time. Incremental investment provides railroad companies with the flexibility to dynamically adjust investment decisions according to observed changes in traffic patterns. Finally, postponed investment can also be beneficial because of cost depreciation and technology advancement.

The goal of the present study was to develop a multiperiod railroad wayside detector location model to maximize the overall benefit from rolling stock coverage. The focus is on a general type of predictive

F. Peng and X. Li, B156 Newmark Civil Engineering Laboratory, and Y. Ouyang, 1209 Newmark Civil Engineering Laboratory, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, MC-250, 205 North Mathews Avenue, Urbana, IL 61801. Corresponding author: Y. Ouyang, yfouyang@illinois.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 148–154.  
DOI: 10.3141/2261-17

wayside detector and the formulation of the problem into a mixed-integer mathematical program. Because of the large scale of real-world applications, it is difficult to use off-the-shelf tools or existing commercial software to solve this model. A range of customized algorithms is thus proposed, including greedy heuristic, interchange heuristic, and Lagrangian relaxation to provide reliable near-optimum solutions in a short time. The mathematical model and solution algorithms are applied to numerical examples, and the influence of multiyear investment on the system design is discussed.

The exposition of this paper is as follows: the next section formulates the mathematic model of the multiperiod wayside detector location problem and proposes several solution techniques. The third section applies the proposed model to a case study and compares performance of the solution methods, and the fourth section briefly discusses how railroad practitioners may potentially use the proposed models in practice. The final section summarizes the paper and discusses future research directions.

## METHODOLOGY

### Model Formulations

The multiperiod wayside detector location problem can be formulated in two ways: a maximum coverage model and a  $K$ -median model. At optimality these two models should yield the same objective value, but solution algorithms to these two models may have different performances.

It is assumed that each railway traffic flow, indexed with  $f$ , is known and deterministic. Let  $F$  be the set of railcar flows and  $L$  the set of candidate locations for detector installation. Let  $T = \{1, 2, \dots, |T|\}$  be the set of time periods in which detectors can be installed. The investment budget (i.e., the maximum number of detectors that can be installed) in period  $t$  is  $r_t$ . Let  $a_{fl}$  denote the relationship between flow  $f \in F$  and location  $l \in L$ ;  $a_{fl} = 1$  if flow  $f$  passes location  $l$ , and  $a_{fl} = 0$  otherwise. If the first installed detector for flow  $f$  is installed in period  $t$ , this flow will be covered since period  $t$ , and its accumulated effectiveness is denoted as  $e_{ft}$ . This value is the sum of the detection effectiveness in all periods from  $t$  to the last period in the analysis horizon. The value of  $|T|$  does not have to be equal to the length of the analysis horizon. For example, if detectors are to be installed in the first 5 years, then  $|T| = 5$ . The analysis horizon could be much longer, for example, 15 years, for which case  $e_{ft}$  would be the total effectiveness from the  $t$ th year to the 15th year. The value of  $e_{ft}$  is nondecreasing over  $t$  but may not be linear because flow  $f$  probably evolves over time rather than remaining at a constant value. The following decision variables are defined:

$\{x_{ft}\}$  = decision variables in the maximum coverage model;  $x_{ft} = 1$  if flow  $f$  is covered by some detector since period  $t$ ;  $x_{ft} = 0$  otherwise;  
 $\{y_{flt}\}$  = decision variables in the  $K$ -median model;  $y_{flt} = 1$  if flow  $f$  is covered by and assigned to a detector at location  $l$  since period  $t$ ;  $y_{flt} = 0$  otherwise; and  
 $\{w_{lt}\}$  = decision variables in both models;  $w_{lt} = 1$  if a detector is installed at location  $l$  in period  $t$ ;  $w_{lt} = 0$  otherwise.

The objective is to determine which detectors are installed in each period in order to maximize the total coverage effectiveness  $z$  across all periods. The maximum coverage model (IP1) is formulated as follows:

$$\max z_{IP1} = \sum_{t \in T} \sum_{f \in F} e_{ft} x_{ft} \quad (1a)$$

subject to

$$\sum_{t \in T} x_{ft} \leq 1 \quad \forall f \in F \quad (1b)$$

$$x_{ft} \leq \sum_{l \in L} a_{fl} w_{lt} \quad \forall f \in F, t \in T \quad (1c)$$

$$\sum_{l \in L} w_{lt} \leq r_t \quad \forall t \in T \quad (1d)$$

$$x_{ft}, w_{lt} \in \{0, 1\} \quad \forall f \in F, l \in L, t \in T \quad (1e)$$

The objective in Equation 1a maximizes the total effectiveness of all flows in all time periods. Constraint 1b ensures that the effectiveness of flow  $f$  is only counted once. Constraint 1c states that flow  $f$  can be covered since time  $t$  only if some detector through which it passes is installed at time  $t$ . Constraint 1d imposes the maximum detector installation number on all periods. Constraint 1e defines the binary variables. In terms of model structure, the traditional maximal-covering model (20) can be extended to IP1 by splitting a single budget constraint into multiple constraints (Constraint 1d) for all periods.

The alternative  $K$ -median model (IP2) is formulated as follows:

$$\max z_{IP2} = \sum_{t \in T} \sum_{l \in L} \sum_{f \in F} a_{fl} e_{ft} y_{flt} \quad (2a)$$

subject to

$$\sum_{t \in T} \sum_{l \in L} y_{flt} \leq 1 \quad \forall f \in F \quad (2b)$$

$$y_{flt} \leq w_{lt} \quad \forall f \in F, l \in L, t \in T \quad (2c)$$

$$\sum_{l \in L} w_{lt} \leq r_t \quad \forall t \in T \quad (2d)$$

$$y_{flt}, w_{lt} \in \{0, 1\} \quad \forall f \in F, l \in L, t \in T \quad (2e)$$

Similar to the maximum coverage model, the objective in Equation 2a maximizes the total effectiveness of all flows in all time periods. Constraints 2b, 2c, 2d, and 2e are equivalent to Constraints 1b, 1c, 1d, and 1e, respectively. Again, the extension from the traditional  $K$ -median model to IP2 on the model structure is that the former has only a single budget constraint (at most  $K$  detectors), but the latter has multiple budget constraints (Constraint 2d). IP1 can be derived from IP2 by letting  $x_{ft}$  equal  $\sum_{l \in L} y_{flt}$  and aggregating Constraint 2c across  $l$ .

Both formulations (IP1 and IP2) yield the same objective value at optimality, and they are NP-hard. In general, no efficient algorithm can obtain exact optimal solutions to large-scale instances of such problems. Heuristics are usually adopted to find nearly optimal solutions. In the following sections, several heuristic methods are proposed and their performances are compared.

### Greedy Algorithm

The greedy algorithm is a simple yet efficient method to solve location problems. Candidate locations that yield the most increase in

inspection effectiveness are added to the selection list one by one until the budget is exhausted. For multiperiod problems, the budget is exhausted in the first period, and then the algorithm moves on to the next period, and so on.

Let  $\bar{L}_t \subseteq L$  be the set of selected locations in period  $t \in T$ , and let  $\bar{F} \subseteq F$  be the set of uncovered flows. Let  $F_l = \{f \in F : a_{fl} = 1\}$  be the set of flows covered by location  $l \in L$ . The greedy algorithm is the same for both the IP1 and IP2 models, as follows:

Step 0.  $\bar{L}_{t'} = \emptyset, \forall t' \in T, \bar{F} = F, t = 1$ .

Step 1. Find

$$l^* = \arg \max_{l \in L} \left\{ \sum_{f \in \bar{F} \cap F_l} e_{fl} \right\}$$

which is the location that yields the highest effectiveness improvement. If  $\bar{L}_t := \bar{L}_t \cup \{l^*\}$ , add this location to the list of selected locations; if  $\bar{F} := \bar{F} \setminus F_{l^*}$ , remove the flows covered by the selected location.

Step 2. If  $|\bar{L}_t| = r_t$  (budget in period  $t$  is exhausted), go to Step 3; otherwise, go to Step 1.

Step 3. If  $t = |T|$ , terminate algorithm; otherwise,  $t := t + 1$  and go to Step 1.

## 1-Interchange Algorithm

The interchange algorithm is a type of local search algorithm that starts from an initial solution and iteratively moves to new neighboring solutions. In each step of the so-called  $k$ -interchange algorithm,  $k$  components of the solution (e.g., detector locations) are perturbed in a search for a better objective value. If the objective value cannot be improved on, the algorithm terminates, and the current solution is called  $k$ -opt.

The 1-interchange algorithm (i.e.,  $k = 1$ ) is widely used for  $K$ -median problems. In one-period  $K$ -median problems, a candidate location only has two possible states: selected or unselected for detector installation. So 1-interchange is usually implemented by simply moving an installed detector to an unselected candidate location. In multiperiod problems, however, a location has  $|T| + 1$  possible states: selected in period  $t = 1, \dots, |T|$  or unselected. 1-Interchange is implemented by exchanging the states of two locations, including the case of exchanging the installation schedules of two detectors that are installed in different time periods.

There are different search strategies for the interchange algorithm. The pair of interchanged detectors–locations can either be the first one that improves the objective value (first improvement) or the one that maximizes the improvement of the objective value (best improvement). Alternatively, the first detector–location can be selected by the first-improvement strategy and the second by the best-improvement strategy. The proposed implementation of the algorithm uses the best-improvement search strategy and targets for a 1-opt solution.

In the proposed 1-interchange algorithm, the candidate neighbors are found as follows:

Step 0. Choose a selected location  $l_1 \in L$  by enumeration.

Step 1. Unselect  $l_1$ . This step will leave a vacancy in the installation time  $t_1 \in T$ .

Step 2. Enumerate all locations that do not have a detector installed in  $t_1$ , and find the one (denoted by  $l_2 \in L$ ) that can improve the objec-

tive value the most by filling the vacancy left by  $l_1$  (i.e., by installing  $l_2$  in  $t_1$ ). In a multiperiod problem, it may happen that  $l_2$  has already been selected in time  $t_2 \neq t_1, t_2 \in T$ . In that case, it will leave another vacancy, and  $l_1$  should be used to fill that vacancy (i.e., install  $l_1$  in  $t_2$ ).

Step 3. Check if a candidate neighbor is obtained.

The proposed 1-interchange algorithm can start from an arbitrary feasible solution, for example, a greedy heuristic solution or a randomly generated solution. However, if it starts from a solution far from optimum (e.g., a random solution), it will usually take a long time to reach 1-opt. If starting from a near-optimum solution (e.g., a greedy heuristic solution), the algorithm will usually terminate much faster. For the present problem, the interchange algorithm can also be used to improve a partial solution, as long as the budget in some time period has not been exhausted. It can be applied in every step of the greedy algorithm whenever a new location has been added to the selection list.

## Lagrangian Relaxation

Lagrangian relaxation is a widely used method to obtain a feasible solution and an optimality gap of mixed-integer programs (21). It can be applied to both IP1 and IP2. For the maximum coverage model, Constraint 1c is relaxed with multipliers  $\mu = \{\mu_{fl}\}$ , where  $\mu_{fl} \geq 0$ . The formulation of the relaxed problem (L1) is as follows:

$$\begin{aligned} z_{L1}^*(\mu) &= \max \sum_{t \in T} \sum_{f \in F} e_{ft} x_{ft} + \sum_{t \in T} \sum_{f \in F} \mu_{ft} \left( \sum_l a_{fl} w_{lt} - x_{ft} \right) \\ &= \max \sum_{t \in T} \sum_{f \in F} (e_{ft} - \mu_{ft}) x_{ft} + \sum_{t \in T} \sum_{l \in L} \left( \sum_{f \in F} \mu_{fl} a_{fl} \right) w_{lt} \end{aligned}$$

subject to Constraints 1b, 1d, and 1e.

The Lagrangian dual (D1) of L1 is as follows:

$$z_{D1}^* = \min_{\mu} z_{L1}^*(\mu)$$

For given  $\mu$  values, the optimum solution  $x_{ft}^*$  and  $w_{lt}^*$  to L1 can be easily determined. For each  $f \in F$ , the  $x_{ft}$  with the largest positive  $e_{ft} - \mu_{ft}$  value is set to 1, or 0 otherwise. For each  $t \in T$ , the  $r_t$  largest positive coefficient  $\sum_{f \in F} \mu_{fl} a_{fl}$  is found and the corresponding variable  $w_{lt}$  is set to 1, or 0 otherwise.

For the  $K$ -median model, Constraint 2b is relaxed with multipliers  $\mu = \{\mu_f\}$ , where  $\mu_f \geq 0$ . The formulation of the relaxed problem (L2) is as follows:

$$\begin{aligned} z_{L2}^*(\mu) &= \max \sum_{t \in T} \sum_{l \in L} \sum_{f \in F} a_{fl} e_{ft} y_{flt} + \sum_{f \in F} \mu_f \left( 1 - \sum_{t \in T} \sum_{l \in L} y_{flt} \right) \\ &= \max \sum_{t \in T} \sum_{l \in L} \sum_{f \in F} (a_{fl} e_{ft} - \mu_f) y_{flt} + \sum_{f \in F} \mu_f \end{aligned}$$

subject to Constraints 2c, 2d, and 2e.

The Lagrangian dual (D2) of L2 is

$$z_{D2}^* = \min_{\mu} z_{L2}^*(\mu)$$

The necessary condition that  $y_{flt} = 1$  in the optimum solution to L2 is that  $a_{fl} e_{ft} - \mu_f > 0$ . Because  $y_{flt} \leq w_{lt}$ , if  $w_{lt} = 1$ , it will contribute to

an objective function as large as  $\sum_{f \in F} (a_{fl} e_{fl} - \mu_f)^+$  (with the definition  $x^+ = \max\{0, x\}$ ). Therefore, for each  $t \in T$ , the  $r_t$  largest  $\sum_{f \in F} (a_{fl} e_{fl} - \mu_f)^+$  is found and the corresponding variable  $w_{lt}$  is set to 1, or 0 otherwise.

It can be proved that D1 and D2 give the same best bound. This is shown in the following proposition:

**Proposition 1.** The optimal objective values of D1 and D2 are equal, that is,  $z_{D1}^* = z_{D2}^*$ .

**Proof.** There still exist integral optimal solutions to L1 and L2, for example, those obtained by the described methods, even if the binary Constraints 1e and 2e are removed. Therefore,  $z_{D1}^*$  and  $z_{D2}^*$  are respectively equal to those obtained by the linear relaxations of IP1 and IP2 (referred to as LP1 and LP2, respectively, i.e.,  $z_{D1}^* = z_{LP1}^*$  and  $z_{D2}^* = z_{LP2}^*$ ) (22). For any feasible solution  $\{x_{fl}\}$  to LP1,  $y_{fl} = a_{fl} w_{lt} \cdot x_{fl} / \sum_{l \in L} a_{fl} w_{lt}$  is a feasible solution to LP2 and gives the same objective value as  $\{x_{fl}\}$  does. For any feasible solution  $\{y_{fl}\}$  to LP2,  $x_{fl} = \sum_{l \in L} a_{fl} y_{fl}$  is a feasible solution to LP1 and also gives the same objective value. Therefore,  $z_{LP1}^* = z_{LP2}^*$ , and thus  $z_{D1}^* = z_{D2}^*$ . ■

The subgradient method can be used to update the multipliers and solve the Lagrangian dual problems (D1 and D2). The initial multipliers can be set as  $\mu_f^0 = 0$  and  $\mu_f^0 = e_{f1}$ , respectively. The step size can be updated in different ways (18); here the well-known approach proposed by Held and Karp is used: the step size is set to two at the beginning, and it is halved if the solution to the dual problem does not improve after three consecutive iteration steps (23).

The solutions to the D1 (or D2) problem may not be feasible to the original problem IP1 (or IP2). Heuristics are applied to construct feasible solutions. In the traditional maximum coverage and  $K$ -median problems, the best locations ( $w_{lt}^*$ ) obtained by solving the relaxed problems are fixed and plugged into the original problems, and their covered customers or flows ( $x_{fl}$  or  $y_{fl}$ ) can then be solved. Cornuejols et al. reported that this method had good performance for  $K$ -median problems (24).

However, in multiperiod problems,  $w_{lt}^*$  of relaxed problems (even if they are feasible to the original problems) may not be reasonable without the additional requirement of

$$\sum_{t \in T} w_{lt} \leq 1 \quad \forall l \in L \quad (3)$$

which ensures that a detector is never installed more than once at the same location. This requirement is redundant for IP1 and IP2 because optimality ensures that Equation 3 is satisfied. However, in L1 and L2, it may be violated by an optimal solution because a duplicated installation may gain extra effectiveness: in L1,  $w_{lt}$  may have a positive coefficient in the objective function; and in L2, the effectiveness of inspecting one flow is allowed to be counted more than once. So it is possible that the optimal solution to the relaxed problem has  $\sum_{t \in T} w_{lt}^* > 1$  for some  $l$ . Such a solution is usually not good for IP1 and IP2 because part of the budget is wasted on duplicated installations. Adding Equation 3 to L1 or L2 will ensure that Equation 3 is satisfied but will also dramatically increase the solution complexity. Thus, an extra heuristic step is proposed to avoid a detector being installed more than once: if  $\sum_{t \in T} w_{lt}^* > 1$  for some  $l$  in the constructed solution, only the one with the smallest  $t$  will be set to 1, and the others will be set to 0. This step removes duplicated installations of detectors and leaves part of the budget unused. The greedy algorithm is then applied to utilize the remaining budget in order to improve the solution. The interchange algorithm should not be used unless the greedy heuristic solution is near optimum (close

to the best bound); otherwise it will take too many steps to reach 1-opt, as explained above.

## CASE STUDY

The data for the case study were obtained from a Class I railroad database that contains over 210,000 units of railcar flows and over 2,000 candidate locations for wayside detector installation. The time horizon for detector effectiveness analysis is 15 years. Because empirical traffic data are only available for 1 year, data for the other 14 years were randomly generated on the basis of information about the first year. If the effectiveness of flow  $f$  is  $e'_{f1}$  in the first year, effectiveness values for the other years ( $e'_{f2}, \dots, e'_{f15}$ ) were randomly generated from a uniform distribution between 0 and  $e'_{f1}$ . The accumulated effectiveness in year  $t$  is calculated as  $e_{ft} = \sum_{i=1}^t e'_{fi}$ . Although the generated data are not necessarily realistic, they are good enough for illustration purposes. The total budget (number of detectors to be installed) in the entire horizon is set to 30, and it is distributed evenly over the first  $|T|$  years.

Commercial software CPLEX (Version 11.1, single thread) was tested for this case study on a 3.0-GHz PC with 2 gigabytes (GB) of RAM, but it failed to give a feasible solution because of the large scale of the problem. Subsequently the following three methods were tried to obtain near-optimum solutions: greedy heuristic, interchange after greedy, and interchange applied in every step of the greedy (i.e., interchange during greedy). The search strategy of the interchange algorithm was best improvement.

The numerical results indicate that spreading the budget over more years will reduce the total effectiveness (Figure 1). For example, as  $|T|$  increases from one to 10, the optimal overall detector surveillance benefit decreases by about 10%. Part of the reason is that a postponed installation loses effectiveness in the years before installation. Figure 1 can help decision makers determine the best number of installations per year to balance cost and effectiveness.

Interchange after greedy is always able to improve the greedy solution for the maximum coverage problem (Figure 2). Interchange during greedy performs better than interchange after greedy in some cases but may obtain a suboptimal solution in others (e.g.,  $|T| = 10$ ).

Lagrangian relaxation was applied to obtain both a best-bound and a feasible solution for the case that  $|T| = 5$ . The algorithm was programmed using C++ and implemented on a 2.0-GHz PC with 1 GB of RAM. To reduce solution time, Lagrangian heuristics were used every 20 iterations to find a feasible solution. The results are shown in Figure 3 and Table 1. Obviously, the  $K$ -median model obtains a better best bound than the maximum coverage model. Table 1 also shows that the  $K$ -median model needs less average CPU time per iteration than the maximum coverage model, so it obtains a better bound within a shorter time.

Although Proposition 1 shows that  $z_{D1}^* = z_{D2}^*$ , IP2 seems to yield a better bound than IP1 in the numerical experiment, probably because the subgradient method does not achieve the optimum value of Lagrangian dual but rather converges to a different value. IP2 also has less average CPU time per iteration than IP1, so it obtains a better bound within a shorter time (Table 1). However, the quality of the best bound depends on the initial multipliers and the step size update approach, and hence no general conclusion about IP1 versus IP2 should be drawn.

Overall, the feasible solutions obtained by Lagrangian relaxation are comparable to those obtained by greedy or greedy-interchange heuristics. For the  $K$ -median problem, the interchange during greedy

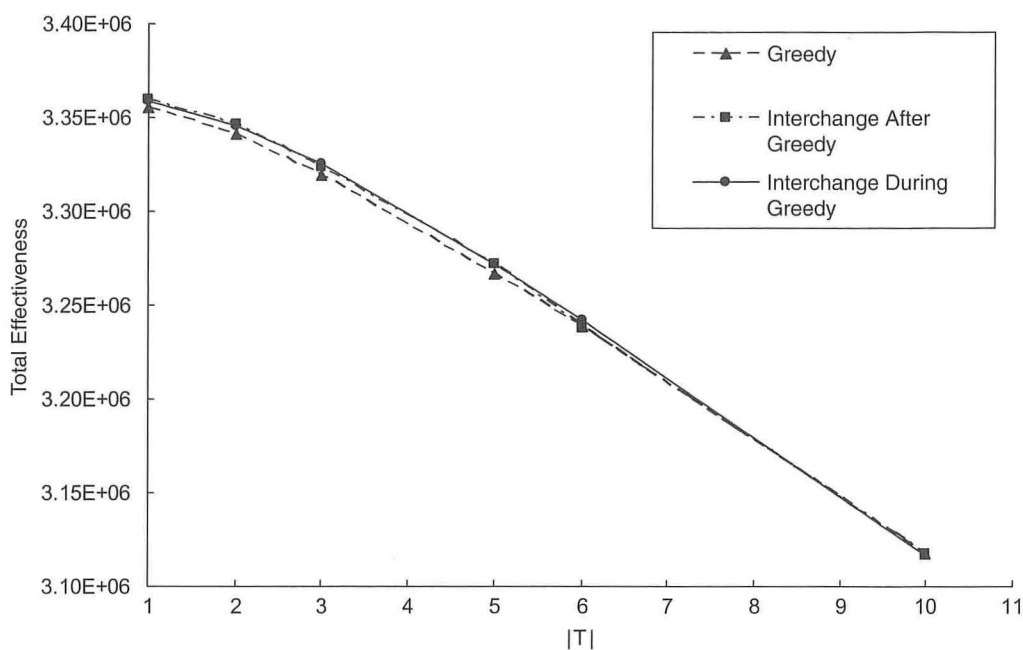


FIGURE 1 Comparison of solutions by different methods.

method obtained an objective value of 3,271,964; Lagrangian relaxation obtained an objective value of 3,185,765. The optimality gaps to the Lagrangian relaxation bounds (from D1 and D2) are 1.86% and 4.45%, respectively. The Lagrangian relaxation algorithm may be further improved if there is a better method to obtain feasible solutions based on information from the relaxed solutions.

#### POTENTIAL USES OF MODEL

The presented models can be used by railroads to determine the best budget allocation plan and the corresponding optimal wayside detector installation locations over multiple years (or other planning periods). For a given budget allocation plan, the model can be applied to obtain the optimal return in terms of maximum inspection

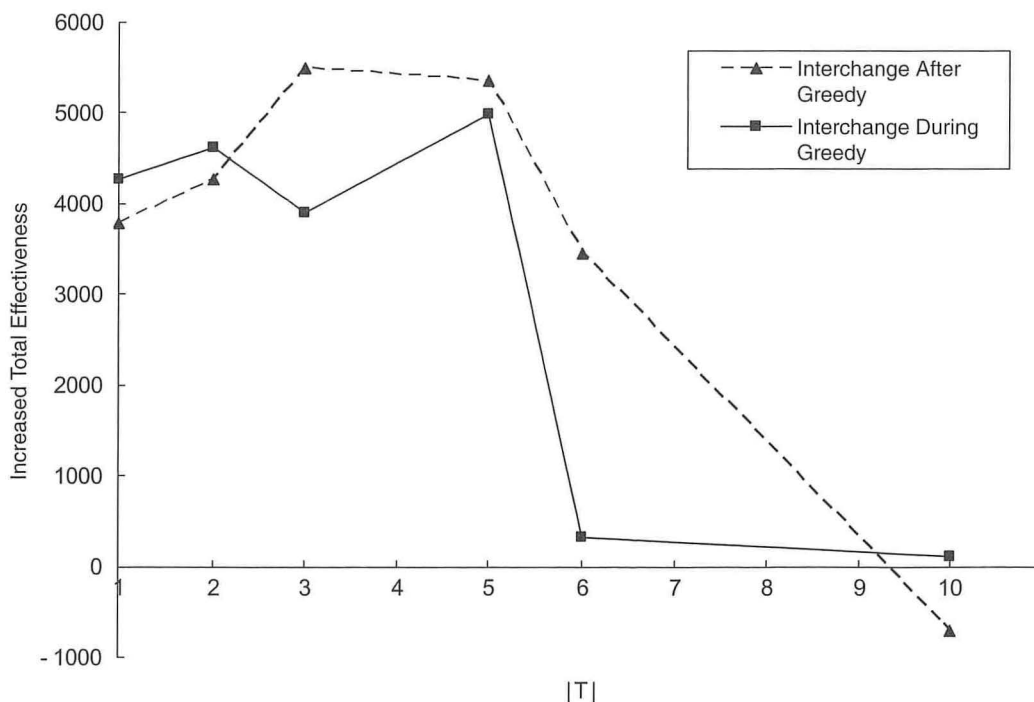


FIGURE 2 Improvements by interchange beyond greedy solutions.

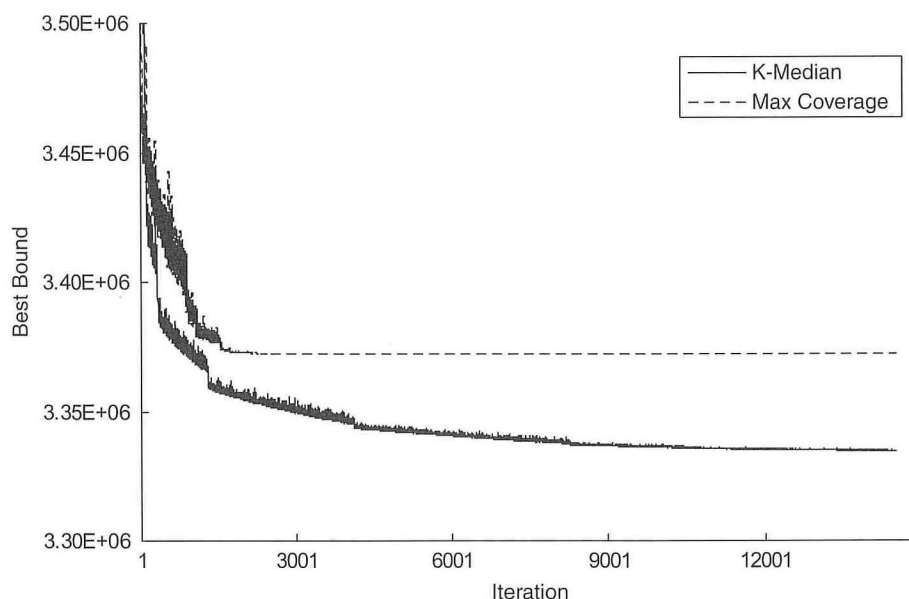


FIGURE 3 Best bounds obtained by Lagrangian relaxation.

effectiveness. The tested allocation plan that yields the maximum optimal return should be chosen for implementation.

If the problem settings (e.g., resource availability and traffic flow routes) change over time during the planning horizon, the original deployment plan may no longer be optimal. In this case, the number of detectors and installation locations should be dynamically adjusted to reflect the changed conditions. Such adjustment can be achieved by developing a slight variation of the proposed model in a rolling horizon framework. The implemented installations can be fixed, the relevant parameters can be changed, and the model can be run to determine the optimal solution for future deployments.

Since various operating uncertainties may have unexpected effects on railcar inspection effectiveness, railroads may be interested in measuring the inspection performance of existing detector systems on a regular basis. If the performance does not meet their expectation, new investment to expand the railcar inspection coverage may be considered. The model proposed in this paper can be used to assess the performance of a given detector system and provide the number and identities of the covered railcars. Such performance assessment will lay the foundation for accurate evaluation of the detector inspection system and cost-effective expansion in the future.

## CONCLUSION AND FURTHER RESEARCH

This paper proposes two formulations of the multiperiod railroad wayside detector installation planning problem. The proposed models extend the wayside detector location design framework to

a more realistic multiyear planning context. Heuristic methods combining greedy and interchange algorithms are used to solve the problem, and Lagrangian relaxation provides a best bound to the objective value. A case study shows that the greedy-interchange heuristics can obtain quite good solutions and are suitable for practical use.

The current models assume that railway traffic flow routes and their volumes are known and deterministic. In real-world practice, the uncertainty of future traffic demand, if significant, will probably affect the incremental planning decision. The proposed model can be extended to address future dynamic and stochastic traffic flow. In some cases, rather than having hard budget constraints, railroads have some leverage in allocating budget to maximize the overall net benefit. In that case, detector installation investment can be incorporated into the optimization objective to decide optimal budget allocations. Technology obsolescence and depreciation should also be considered in budget allocation adjustments. In some applications, more frequent detections generate higher benefits. The proposed model may be extended to address the extra coverage benefits from repeated detections. In systems with multiple types of detectors, the proposed model can be applied to each type of detector separately to obtain the overall optimal detector installation design if the effectiveness measures and budget allocations are independent across detector types; otherwise, the model should be adapted as discussed by Ouyang et al. (11). The proposed model can also be extended to highway traffic surveillance applications. Although a highway surveillance system mainly focuses on macroscopic traffic statistics (e.g., flow, density, and speed), while a railroad monitoring system focuses on the microscopic health conditions of individual railcars, they share fundamental similarities: the objective is to inspect as much traffic as the budget allows (or to optimize the trade-off between inspection effectiveness and detector investment), and the challenge arises from the complex interrelationship among candidate installation locations that share network traffic flows. Hence, the proposed modeling framework and solution techniques could be adapted to solve multiperiod location design for highway surveillance sensors.

TABLE 1 Comparison of Lagrangian Relaxation Results ( $|I| = 5$ )

Result	Model	
	Maximum Coverage	<i>K</i> -Median
Best feasible solution (larger the better)	3,139,738	3,185,765
Best-bound solution (smaller the better)	3,333,985	3,371,950
Average CPU time per iteration step (s)	19.7	17.6

## ACKNOWLEDGMENTS

This research was supported in part by the U.S. National Science Foundation. The authors thank CSX Transportation, Inc., for data support and the anonymous reviewers for their valuable comments.

## REFERENCES

1. Resor, R. R., and A. M. Zarembski. Factors Determining the Economics of Wayside Defect Detectors. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.
2. Post, W. M. Protective System for Railways. U.S. Patent 2063336, 1936.
3. Post, W. M. Protective System for Railways. U.S. Patent 2095616, 1937.
4. Burpee, C. M. (ed.). *Railway Engineering and Maintenance Cyclopedia*, 6th ed. Simmons-Boardman Publishing Company, Chicago, Ill., 1945.
5. Austin, K. B. Hotbox Signal for Railway Trains. U.S. Patent 2486546, 1949.
6. Gallagher, C. A., and W. M. Pelino. Hot-Box Detector. U.S. Patent 2880309, 1959.
7. Lagnebäck, R. *Evaluation of Wayside Condition Monitoring Technologies for Condition-Based Maintenance of Railway Vehicles*. Licentiate thesis. Luleå University of Technology, Luleå, Sweden, 2007. <http://publ.ltu.se/1402-1757/2007/18/LTU-LIC-0718-SE.pdf>. Accessed January 2008.
8. Southern, C., D. Rennison, and U. Kopke. RailBAM: An Advanced Bearing Acoustic Monitor: Initial Operational Performance Results. Presented at Railway Engineering Conference, Darwin, Australia, 2004. [http://www.railbam.com.au/railbam/railbam\\_o\\_performance.pdf](http://www.railbam.com.au/railbam/railbam_o_performance.pdf).
9. Wolf, G. P., and F. J. Peterson. Application and Validation of the Automated Truck Performance Measurement System. *Proc., IEEE/ASME Joint Railroad Conference 1998*, Piscataway, N.J., 1998, pp. 127–134.
10. Tournay, H., T. Wolgram, S. Cummings, and S. Chapman. *Interpreting Truck Performance Detector Data to Establish Car and Truck Condition*. R-977. Transportation Technology Center, Inc., Association of American Railroads, Pueblo, Colo., 2006.
11. Ouyang, Y., X. Li, C. Barkan, A. Kawprasert, and Y.-C. Lai. Optimal Locations of Railroad Wayside Defect Detection Installations. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 24, No. 5, 2009, pp. 309–319.
12. Gribbon, A. P. Field Test of Nonintrusive Traffic Detection Technologies. *Mathematical and Computer Modelling*, Vol. 27, No. 9–11, 1998, pp. 349–352.
13. Yang, H., Y. Iida, and T. Sasaki. An Analysis of the Reliability of an Origin–Destination Trip Matrix Estimated from Traffic Counts. *Transportation Research Part B*, Vol. 25, No. 5, 1991, pp. 351–363.
14. Yang, H., and J. Zhou. Optimal Traffic Counting Locations for Origin–Destination Matrix Estimation. *Transportation Research Part B*, Vol. 32, No. 2, 1998, pp. 109–126.
15. Ehlert, A., M. G. Bell, and S. Grosso. The Optimisation of Traffic Count Locations in Road Networks. *Transportation Research Part B*, Vol. 40, No. 6, 2006, pp. 460–479.
16. Fei, X., H. S. Mahmassani, and S. M. Eisenman. Sensor Coverage and Location for Real-Time Traffic Prediction in Large-Scale Networks. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2039, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 1–15.
17. Fei, X., and H. S. Mahmassani. Two-Stage Stochastic Model for Sensor Location Problem in a Large-Scale Network. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
18. Li, X., and Y. Ouyang. Reliable Traffic Sensor Deployment Under Probabilistic Disruptions and Generalized Surveillance Effectiveness Measures. Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.
19. Li, X., and Y. Ouyang. Reliable Sensor Deployment for Network Traffic Surveillance. *Transportation Research Part B*, Vol. 45, No. 1, pp. 218–231.
20. Daskin, M. S. *Network and Discrete Location: Models, Algorithms and Applications*, John Wiley & Sons, Inc., New York, 1995.
21. Fisher, M. L. The Lagrangian Relaxation Method for Solving Integer Programming Problems. *Management Science*, Vol. 27, No. 1, 1981, pp. 1–18.
22. Geoffrion, A. M. Lagrangian Relaxation for Integer Programming. *Mathematical Programming Study*, Vol. 2, 1974, pp. 82–114.
23. Held, M., and R. M. Karp. The Traveling Salesman Problem and Minimum Spanning Trees: Part II. *Mathematical Programming*, Vol. 1, 1971, pp. 6–25.
24. Cornuejols, G., M. L. Fisher, and G. L. Nemhauser. Location of Bank Accounts to Optimize Float: An Analytic Study of Exact and Approximate Algorithms. *Management Science*, Vol. 23, No. 8, 1977, pp. 789–810.

---

*The Railroad Operating Technologies Committee peer-reviewed this paper.*

# Track Maintenance of Heavy Haul Railways with Resilient Versus Stiff Fastenings

K. Giannakos

The average stress on the tie-ballast contact surface is used to examine the stress on the seating of the track. According to AASHTO testing for road construction, increasing the pressure on the ballast by 10% results in a decrease that is 1.3 to 1.5 times more rapid in the geometry of the track and a corresponding increase in the maintenance cost. The type of fastening used, rigid or resilient, plays a major role in the magnitude of the stress at the seating surface of the tie on the ballast. Track maintenance and renewal are planned accordingly, always taking into consideration local conditions, on the basis of a combination of control data from measuring systems, visual observation, and economic data. This paper presents a parametric investigation of the influence of rigid versus resilient fastenings on track maintenance of heavy haul railroads.

The railway track is a multilayered structure consisting of a vertical succession of various materials or layers of materials that define the final position of the rail running table as well as the properties of the track itself as it reacts to the action created from the motion of the railway vehicle. Each material or layer that constitutes the line can be simulated by a combination of a spring with spring constant  $k_i$  and a damper with damping coefficient  $c_i$ . The weak links in this multilayered construction are the ballast and the substructure. These develop residual deformations, deflections, and lateral displacements proportionally related to the deterioration of the geometry of the track, also referred to as the quality of the track.

Since the permanent deformation of the track is a percentage of the total deflection resulting from the train circulation, it is imperative to reduce the deflection of the layers as much as possible as well as the lateral displacement. According to the theoretical analysis of Winkler (1), Weaver et al. (2), and Zimmermann (3), the deflection of the track as an infinite beam on an elastic foundation should be high enough to distribute the acting load to a longer section of the track and thus to reduce the reacting force at each point. This amount of deflection can be provided by a resilient fastening and its rail pad.

Generally, the ballast must ensure damping of the larger part of the vibrations of the train, adequate load distribution, and prompt drainage of rainwater. The ballast and substructure are the elements of the track that develop residual deformations directly connected to the deterioration of the geometry of the track. The smaller the

residual deformations and their increase are over time, the better the quality of the track is.

AASHTO's road construction equation for maintenance costs is also applicable for a railway track (4, 5):

decrease in track geometry quality

$$= (\text{increase in stress on the ballast bed})^m$$

where  $m = 3$  to 4.

The decrease in track geometry quality has a proportionate effect on maintenance costs and is related to the stresses on the ballast bed and the degree of fouling of the ballast bed. The latter negatively influences the preservation of track geometry. Since stress is equal to the ratio of the actions on the sleeper (reaction per point) to the seating surface of the tie, and the seating surface of each tie type is standard, a more precise estimation of the actions on the track should be performed in order to calculate the stresses (and consequently the maintenance cost) on the ballast bed. In a related study, the author recently presented a relationship between ballast hardness and fouling to determine the influence of ballast quality on maintenance cost (for preserving track geometry) (6).

To meet the needs presented above, a new approach for the actions on sleepers and consequently on the ballast bed and subgrade was developed that takes into account the real conditions of the track (e.g., maintenance) (4, 7). This approach led to the adoption of more strict specifications for railway track structural elements in Greece, especially for very resilient fastenings, ballast hardness, the specifications for substructure construction, and so forth. In the present paper the Giannakos method is described and applied to calculate the actions on the track and the stress on the ballast bed for heavy haul railways in the case of stiff and very resilient fastenings as laid on track in the United States (4). The results are compared with the American Railway Engineering and Maintenance-of-Way Association (AREMA) method.

## TRACK MAINTENANCE AND VEHICLE-TRACK SYSTEM

Track analysis of the actions on the track and the track's constituent components have led to the examination of the ensemble vehicle-track system, which consists of the vehicle with all its suspensions (primary and secondary), the car body and all the rolling instruments (axles or bogies), and the track, which comprises the rail-sleeper-fastenings-ballast substructure. A subsystem of the latter system constituting the railway track is the rail-sleeper with its fastenings

Department of Civil Engineering, University of Thessaly, Volos, 108 Neoreion Str., Piraeus 18534, Greece. k.giannakos@ontelecoms.gr.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 155–162.  
DOI: 10.3141/2261-18

and ballast, which, in classic terminology, is the superstructure of the track (today, the blanket layer and subballast are included in the superstructure).

The whole system generally consists of five parameters (Figure 1) of spring constant  $\rho_i$ . If any of the five is changed, the rest are affected, and the whole system reacts differently. During the planning of a high-speed line, all the details and specifications for the five parameters should be followed—with no exceptions—as otherwise significant problems to the overall functioning of the system may arise, with serious consequences either to the safety of traffic or to the cost of operation and maintenance of the track, or both. Track maintenance is affected by the loading conditions of the vehicle motion combined with the constitutive layers of the track and their stiffness coefficients. This combination determines the acting reaction—action on each discrete point of support (each tie) of the rail and the implied stresses on the ballast bed and substructure. The stress on the ballast bed is defined as the ratio of the acting reaction divided by the seating surface of the tie. In the international literature, the average stress on the sleeper—ballast contact surface is used to examine the stress on the seating of the track.

When the pressure on the ballast is increased, maintenance costs also increase as a result of the deterioration of the track geometry. In dimensioning as well as in the selection of the individual materials comprising a railway track, the weak links are the ballast and substructure. These two elements present residual deformations in the form of subsidence and lateral displacement and are directly connected to the deterioration of the so-called geometry of the track, which can be described more specifically as the quality of the track. The smaller the residual deformations and the slower their fluctuation over time are, the better the quality of the track is. It is therefore imperative to reduce the development of subsidence and lateral displacement as much as possible. For a given quality of ballast material, as far as the part of the deformation caused by the ballast and the earthworks of the track are concerned, a reduction in subsidence is accomplished by the correct combination and use of heavy track machinery (e.g., ballast regulator, tamping machine, dynamic stabilizer). For the layers underneath the ballast a well-executed construction is required: crushed stone material in the upper layer and 100% modified Proctor or 105% Proctor normal compaction. This specification implies an almost undeflected substructure. Taking into account the need of significant deflection to achieve a distribution of the acting load to the adjacent ties, this deflection should be offered by the fastening and its rail pad. As for ballast fatigue, the lit-

erature assumes a uniform distribution of stresses under the sleeper and without further details uses the mean value of pressure.

It is obvious from the analysis above that the type of fastening used plays a major role in the magnitude of the stress on the contact (seating) surface between the tie and the ballast bed. Track maintenance and renewal are planned accordingly, always taking into consideration local conditions, based on a combination of control data from measuring systems, visual observation, and economic data. Like all construction, each track has a predetermined life cycle during which maintenance work is necessary for the provision of the basic, minimum standards of quality and safety. For a conventional superstructure consisting of rail, fastenings, sleepers, and ballast, an optimum life cycle exists from an economic point of view, dependent on the daily traffic and annual tonnage.

## AREMA AND GIANNAKOS METHODS FOR ACTIONS ON TRACKS

The stiffness (elasticity) coefficient  $\rho_i$  (in kilonewtons per millimeter) of each layer of the track structure is a spring coefficient (as in Hooke's law) that contributes to the total track stiffness coefficient  $\rho_{\text{total}}$ . The calculation of stiffness  $\rho_i$  and  $\rho_{\text{total}}$  is used to determine the action and reaction on a sleeper. After experimental on-the-track investigation, Eisenmann and Mattner reported that the theoretical subsidence is the same as that derived from the calculation of the track's vertical stiffness (8). Eisenmann also ascertained that theoretical calculations performed for the dimensioning of the superstructure correspond to the average value of the measurements (9, 10). The calculation of  $\rho_{\text{total}}$  is performed for springs in a parallel arrangement by

$$\frac{1}{\rho_{\text{total}}} = \frac{1}{\rho_{\text{rail}}} + \frac{1}{\rho_{\text{pad}}} + \frac{1}{\rho_{\text{sleeper}}} + \frac{1}{\rho_{\text{ballast}}} + \frac{1}{\rho_{\text{substructure}}} \quad (1)$$

### AREMA Method

The influence curve for deflection  $w$  given in AREMA (11, p. 16-10-29) is used to determine the largest value of the stress  $p_{\text{max}}$ , and the maximum reaction on rail seat  $R_{\text{max}}$  on an individual tie is given by

$$R_{\text{max}} = p_{\text{max}} \cdot \ell = k \cdot w_{\text{max}} \cdot \ell = k \cdot \frac{\beta \cdot Q_{\text{total}}}{2 \cdot k} \cdot \ell \quad (2)$$

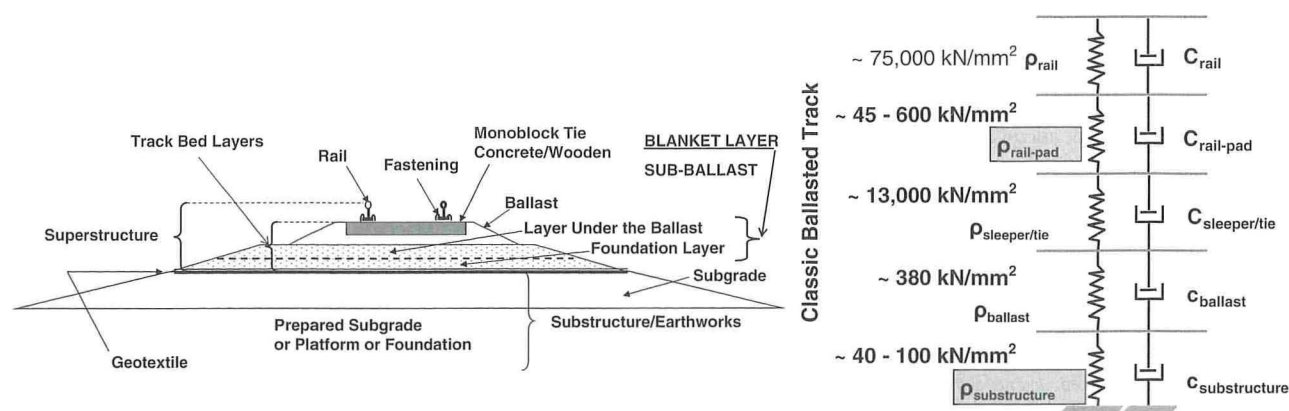


FIGURE 1 Typical cross section of ballasted track and simulation as combination of springs and dampers.

where

$$\begin{aligned} p_{\max} &= \text{rail base stress per unit of track length} \\ &\quad (\text{lb/in or kN/mm}), \\ w_{\max} &= \text{maximum deflection,} \\ \ell &= \text{distance between ties, and} \\ k &(\text{lb/in.}^2 \text{ or kN/mm}^2) = \rho_{\text{total}}/\ell. \end{aligned}$$

$Q_{\text{total}}$  is derived in Equation 5, and  $\beta$  is given by

$$\beta = \sqrt[4]{\frac{k}{4 \cdot E \cdot I}} = \sqrt[4]{\frac{\rho_{\text{total}}}{4 \cdot E \cdot I \cdot \ell}} \quad (3)$$

where  $E$  is the modulus of elasticity, and  $I$  is the moment of inertia of the rail. Equation 2 is transformed into

$$R_{\max} = \sqrt[4]{\frac{\rho}{4E \cdot I \cdot \ell}} \cdot \frac{Q_{\text{total}} \cdot \ell}{2} = \frac{1}{2\sqrt{2}} \cdot \sqrt[4]{\frac{\rho \cdot \ell^3}{E \cdot I}} \cdot Q_{\text{total}} = \bar{A}_{\text{stat}} \cdot Q_{\text{total}} \quad (4)$$

where  $\bar{A}_{\text{stat}}$  is the static coefficient of the sleeper's reaction and is given in Equation 6, as described by Giannakos (12).

$$Q_{\text{total}} = Q_{\text{wheel}} \cdot (1 + \theta) \quad \theta = \frac{D_{33} \cdot V}{D_{\text{wheel}} \cdot 100} \quad \text{IF} = (1 + \theta) \quad (5)$$

where

$$\begin{aligned} D_{33} &= \text{diameter of a wheel of 33 in.,} \\ D_{\text{wheel}} &= \text{diameter of the wheel (in.) of vehicle examined,} \\ V &= \text{speed (mph), and} \\ \text{IF} &= \text{impact factor.} \end{aligned}$$

$$\bar{A}_{\text{stat}} = \frac{1}{2\sqrt{2}} \cdot \sqrt[4]{\frac{\ell^3 \cdot \rho_{\text{total}}}{E \cdot I}} \quad (6)$$

The influence line of deflection  $w$  and reaction–action  $R$  on each tie or fixation point of the rail on the track's superstructure is given by

$$\eta(x) = e^{-\beta x} \cdot [\cos(\beta x) + \sin(\beta x)] \quad (7)$$

Finally, according to AREMA, reaction  $R$  is given by the following equation as derived from Equation 2:

$$R = \frac{Q_{\text{wheel}} \cdot \beta \cdot \ell}{2} \cdot \eta(x) \Rightarrow R_{\max} = \bar{A}_{\text{stat}} \cdot Q_{\text{wheel}} \quad (8)$$

### Giannakos Method

Giannakos presented a method verifying the results derived from an examination of the cracking appearance of concrete ties laid on track (4, 7). The reactions–actions on the track panel, per sleeper, are calculated with the following equation:

$$R_{\max} = (Q_{\text{wh}} + Q_{\alpha}) \cdot \bar{A}_{\text{dynam}} + \lambda \cdot \sqrt{\sigma(\Delta Q_{\text{NSM}})^2 + \sigma(\Delta Q_{\text{SM}})^2} \quad (9)$$

where

$$\begin{aligned} Q_{\text{wh}} &= \text{static wheel load,} \\ Q_{\alpha} &= \text{load caused by cant deficiency (super-} \\ &\quad \text{elevation deficiency),} \\ \bar{A}_{\text{dynam}} &= \text{dynamic coefficient of sleeper's} \\ &\quad \text{reaction,} \\ \lambda &= \text{coefficient of dynamic load (3 for a} \\ &\quad \text{99.7\% probability of appearance), and} \\ \sigma(\Delta Q_{\text{NSM}}) \text{ and } \sigma(\Delta Q_{\text{SM}}) &= \text{standard deviation of the dynamic load} \\ &\quad \text{from the nonsuspended mass } (m_{\text{NSM}}) \\ &\quad \text{and suspended mass } (m_{\text{SM}}) \text{ of each axle,} \\ &\quad \text{respectively.} \end{aligned}$$

Then

$$\begin{aligned} \sigma(\Delta Q_{\text{NSM}}) &= m_{\text{NSM}} \cdot \sigma(\gamma) = \frac{k_{\text{rail}} \cdot m_{\text{NSM}} \cdot V}{1,000} \\ \sigma(\Delta Q_{\text{SM}}) &= \frac{V - 40}{1,000} \cdot N_L \cdot Q_{\text{wheel}} \end{aligned} \quad (10)$$

where  $k_{\text{rail}}$  is the coefficient of the condition of the running rail table and  $N_L$  is a coefficient normally equal to 1 to 1.2.

Then

$$\bar{A}_{\text{dynam}} = \frac{1}{2\sqrt{2}} \cdot \sqrt[4]{\frac{\ell^3 \cdot h_{\text{TR}}}{E \cdot I}} \quad (11a)$$

where  $h_{\text{TR}}$  is the total dynamic stiffness of the track, which is given by

$$h_{\text{TR}} = \rho_{\text{dynam}} = \frac{1}{2 \cdot \sqrt{2}} \cdot \sqrt[4]{E \cdot I \cdot \frac{\rho_{\text{total}}}{\ell}} \quad (11b)$$

The results from this method predict an extended appearance of cracks in twin-block sleepers that was indeed observed on the Greek permanent way (60% or even higher) (described below). For the estimation of the service load, the pad stiffness is calculated through a trial-and-error procedure that ensures equilibrium among the numerous spring components of the track system. This method has been adopted for a precast concrete railway track system design (13).

For more details about these methods as well as about the methods in the German and French literature the interested reader is referred to Giannakos (12).

## CALCULATION OF STRESSES ON TIE-BALLAST CONTACT SURFACE

### AREMA Method

According to AREMA (11, pp. 16-10-31–16-10-32), the mean stress  $\bar{p}$  on the ballast bed at the seating surface of the tie is

$$\bar{p} = \frac{(R_{\max})}{\frac{1}{3} \cdot A_b} = \frac{(\bar{A}_{\text{stat}} \cdot Q_{\text{total}})}{\frac{A_b}{3}} \quad (12)$$

where  $Q_{\text{total}}$  is given by Equation 5, and  $A_b$  is the seating surface of the tie.

Similarly, according to AREMA (11, p. 30-4-5) the mean stress  $\bar{p}_1$  on the ballast bed at the seating surface of the tie is

$$\begin{aligned}\bar{p}_1 &= \frac{(2 \cdot Q_{\text{wheel}}) \cdot \left(1 + \frac{\text{IF}}{100}\right) \cdot \left(\frac{\text{DF}}{100}\right)}{A_b} \\ &= \frac{(2 \cdot Q_{\text{wheel}}) \cdot \left(1 + \frac{\text{IF}}{100}\right) \cdot \bar{A}_{\text{stat}}}{A_b}\end{aligned}\quad (13a)$$

where

$(2 \cdot Q_{\text{wheel}})$  = load of the two wheels of the axle,  
IF = impact factor coefficient equal to  $\theta \cdot 100$ , and  
DF = distribution factor equal to  $100 \cdot \bar{A}_{\text{stat}}$ .

Equation 13a is transformed into Equation 13b for one wheel, where  $\bar{p}_1 = \bar{p}_2$ :

$$\bar{p}_2 = \frac{Q_{\text{wheel}} \cdot \left(1 + \frac{\text{IF}}{100}\right) \cdot \bar{A}_{\text{stat}}}{\frac{A_b}{2}} = \frac{Q_{\text{total}} \cdot \bar{A}_{\text{stat}}}{\frac{A_b}{2}} = \bar{p}_1 \quad (13b)$$

Equations 12 and 13 do not yield the same result, since Equation 13b considers the middle third of the tie unloaded. This outcome is dependent on the condition of the track (well maintained or not), but in any case the effective length ( $L_{\text{eff}}$ ) of a tie of length  $L_{\text{tie}}$  in a track of gage  $e$  is per rail seat:

$$L_{\text{eff}} = (L_{\text{tie}} - e) \quad (14)$$

If  $L_{\text{tie}} = 2.60$  m and  $e \approx 1.50$  m, then per rail  $L_{\text{eff}} = 1,100$  mm and not 867 mm (i.e.,  $2,600 \div 3$ ) as in AREMA (11, pp. 16-10-31–16-10-32), and Equation 12 is proposed. In the present paper Equations 12 and 15 are used ( $L_{\text{eff}}$  is calculated from Equation 14):

$$\bar{p}_{\text{ballast}} = \frac{Q_{\text{wheel}} \cdot \left(1 + \frac{\text{IF}}{100}\right) \cdot \bar{A}_{\text{stat}}}{L_{\text{eff-tie}} \cdot b_{\text{tie}}} = \frac{Q_{\text{total}} \cdot \bar{A}_{\text{stat}}}{(L_{\text{tie}} - e) \cdot b_{\text{tie}}} \quad (15)$$

where

$\bar{p}_{\text{ballast}}$  = mean stress on the ballast bed as if the distribution of pressures were uniform,  
 $L_{\text{tie}}$  = length of the tie (i.e., 8 ft 6 in. [2,590 mm]),  
 $e$  = gage of the track ( $\sim 1,500$  mm), and  
 $b_{\text{tie}}$  = width of the tie (in.) at the seating surface.

Normally for concrete ties  $b_{\text{tie}}$  fluctuates from 8 to 13 in. (11, p. 30-4-18), with more representative values at 12 in. However, since for wooden ties the normal width is 9 in. (11, p. 30-3-4), for comparison reasons in this paper 9 in. (228.6 mm) is used for the width of both wooden and concrete ties.

### Giannakos Method

For the blanket layers, subgrade, and prepared subgrade (Figure 1) that constitute the formation, dimensioning is performed with design loads and actions as derived by Equation 9 with two times (or one time

for the upper surface of the prepared subgrade) the standard deviation of the dynamic component of the load instead of three, corresponding to a possibility of 95.5% instead of 99.7% for the earthworks (4, 7, 12). Thus the following equation is derived from Equation 9:

$$R = \bar{A}_{\text{dynam}} \cdot (Q_{\text{wheel}} + Q_{\alpha}) + \left(2 \cdot \sqrt{[\sigma^2(\Delta Q_{\text{NSM}})]^2 + [\sigma^2(\Delta Q_{\text{SM}})]^2}\right) \quad (16)$$

The average pressure on the upper surface of the formation (ballast) can be calculated by the following equations:

$$\bar{p} = \bar{A}_{\text{subsidence}} \cdot (Q_{\text{wheel}} + Q_{\alpha}) + \frac{\sqrt{[\sigma(\Delta Q_{\text{NSM}})]^2 + [\sigma(\Delta Q_{\text{SM}})]^2}}{h_{\text{TR}}} \cdot C \quad (17)$$

$$C = \frac{P_{\text{total}}}{\left(\frac{A_b}{2}\right)} \quad (18)$$

$$\bar{A}_{\text{subsidence}} = \frac{1}{2\sqrt{2}} \cdot \sqrt{\frac{\ell^3}{E \cdot I \cdot h_{\text{TR}}^3}} \quad (19)$$

where  $C$  is the ballast coefficient (N/mm<sup>3</sup>) equal to the quotient of stiffness divided by the active seating surface of the semitie,  $A_b$  is the tie's seating surface (for monoblock sleepers the central non-loaded area should be subtracted), and the rest of the parameters are as defined above.

The subsidence  $w_{\text{total}}$  of the track multilayered structure should be calculated by

$$w_{\text{total}} = \bar{A}_{\text{subsidence}} \cdot (Q_{\text{wheel}} + Q_{\alpha}) + \frac{2 \left( \sqrt{[\sigma(\Delta Q_{\text{NSM}})]^2 + [\sigma(\Delta Q_{\text{SM}})]^2} \right)}{h_{\text{TR}}} \quad (20)$$

The literature cites that measurements on the track indicate that dispersion should be taken into consideration, which enters the theoretical calculation through coefficients depending on the probability of the appearance of various individual parameters (9, 10). In the German literature, a smaller coefficient of probability of appearance (95.5% with  $t = 2$  or even 68.3% with  $t = 1$ ) is used for the formation of the track (10).

### APPLICATION FOR HEAVY HAUL RAILROAD TRACK WITH STIFF AND RESILIENT FASTENINGS

The heavy haul freight railway transportation specifications in the United States are wheel load 17.69 tonnes (t) (39,006 lb; 35.38 t per axle); maximum speed 60 mph (96.6 km/h or approximately 100 km/h); and distance between two consecutive ties 24 in. (60.96 cm). Ahlbeck et al. (14) and Hay (15) propose values in the order of 4,450 lbs or 2,018 t per wheel for the unsprung (nonsuspended) mass. In this paper this value is used for the unsprung mass.

The use of wooden ties equipped with stiff fastenings with no rail pad, like the elastic spike (Figure 2a) or similar cut spike with tie plate (Figure 2b) on the heavy haul railroad track, should be compared

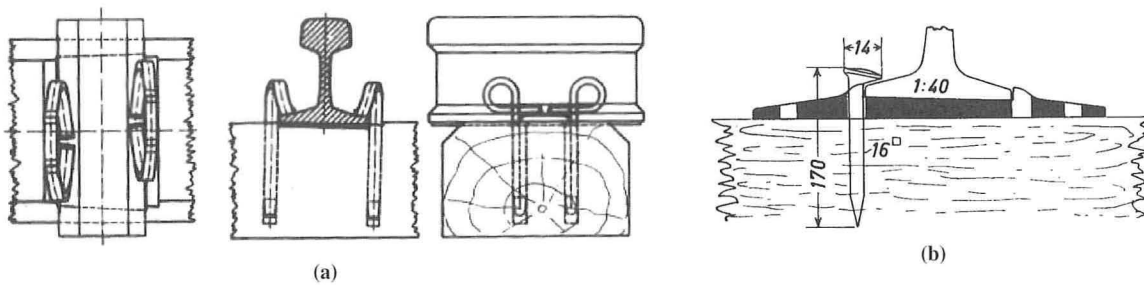


FIGURE 2 Rigid fastening of rail on wooden tie without rail pad: (a) elastic spike and (b) cut spike with tie plate as used in the United States.

with a very resilient fastening. The spring coefficient ( $\rho_{\text{pad}}$ ), which contributes almost 50% of the total static track stiffness coefficient  $\rho_{\text{total}}$  (Equation 1), does not exist in these cases. The adoption of high-resilience fastenings, such as the W24 with clip SkI-24 and highly resilient strain-attenuating Zw700WIC rail pads (Figure 3), which can be adopted for switches (and plain track also), would help decisively to reduce stresses and consequently to prolong the life cycle and decrease the maintenance costs of the track.

To calculate the real acting forces on the superstructure and the ties, the exact rigidity of the elastic pad of the fastening for each combination of parameters must be determined by applying the equations above in a multilayered construction with polyparametrical functions. In the case of the W24 fastening, the tie pad stiffness of the Zw700WIC pad must be calculated and used according to its load-deflection curve, taking into account—the toe-load that results from the fastening clip's preload (12). The load-deflection curve is characteristic and unequivocally exclusive of each element. The same type of product (e.g., pad Zw700) belonging to different brands presents quite different load-deflection curves (as well as life cycle) for the same specifications, meeting the same minimum demands. The most adverse curve is used because it describes the behavior of the pad during the approach of the wheel, since the second curve describes the unloading of the pad after the removal of the wheel. The stiffness of the substructure varies from 40 kN/mm for muddy substructure to 250 kN/mm for rocky tunnel bottom with not enough ballast thickness. Each time this stiffness changes in Equation 1, the acting stiffness of the tie pad also changes.

So the method included in the regulations for calculating the pad stiffness from two discrete load values (i.e., 18 and 70 kN) does not accurately describe the real situation, in which an equilibrium occurs among the various springs that comprise the track system. A trial-and-error method must be used to more accurately estimate the stiff-

ness of the pad, in each case taking into account the toe-load that results from the fastening clip's preload. In this paper the stiffness of the pad is calculated with the trial-and-error method (for the load-deflection curve), and then the acting-forces loads on the ties with resilient fastenings are calculated. For the W24 fastening with SkI-24 tension clamp and the soft Zw700WIC pad for heavy haul rail transport, the trial-and-error method is used to find the equilibrium for the total number of layers (springs). For the rigid fastening there is no trial-and-error method;  $\rho_{\text{total}}$  is calculated directly from Equation 1.

Applying the AREMA method (Equation 8) and Giannakos method (4, 12) gives the values of the actions on the ties shown in Figure 4. These values are derived for the parameters above and for 69.25 kg/m rail type 140 RE (11, 16); concrete tie approximately 2.59 to 2.60 m (363 kg); and a rail running table coefficient for two conditions ( $k_{\text{rail}} = 9$  and  $k_{\text{rail}} = 12$ ), that is, for an average and a worse situation of the rail. Additional details on  $k_{\text{rail}}$  are available (7, 12).

In this paper, the rigid fastening without pad (either cut spike or elastic spike;  $\rho_{\text{pad}} = 75,000$  kN/mm) is used in combination with the more flexible wooden ties, presenting a stiffness coefficient of  $\rho_{\text{tie}} = 800$  kN/mm. The resilient fastening ( $85.18$  kN/mm  $\leq \rho_{\text{pad}} \leq 91.41$  kN/mm) is used in combination with the stiffer concrete tie, presenting a stiffness coefficient of  $\rho_{\text{tie}} = 13,500$  kN/mm. This comparison is done from a perspective more favorable for the rigid fastening since the more flexible wooden tie contributes to a longer distribution of load along the railway track in comparison to the stiffer concrete tie. The fastening stiffness ( $\rho_{\text{pad}} = 75,000$  kN/mm) is derived from the steel base plate only, without any contribution from even a very stiff rail pad. The substructure stiffness coefficient fluctuates from 40 kN/mm for pebbly subgrade to 250 kN/mm for a rocky tunnel bottom with a small ballast cross-section height.

Finally, the W24 resilient fastening on concrete ties gives very good results for heavy haul rail transport compared with the rigid

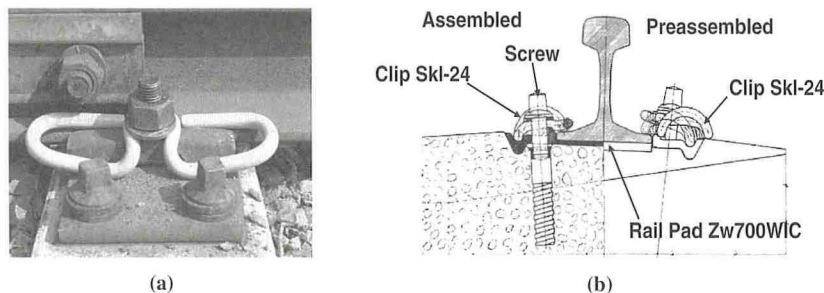


FIGURE 3 Very resilient fastening W24 with clip SkI-24 and rail pad Zw700WIC (a) in track photograph and (b) on concrete tie in (left) assembled and (right) preassembled position.

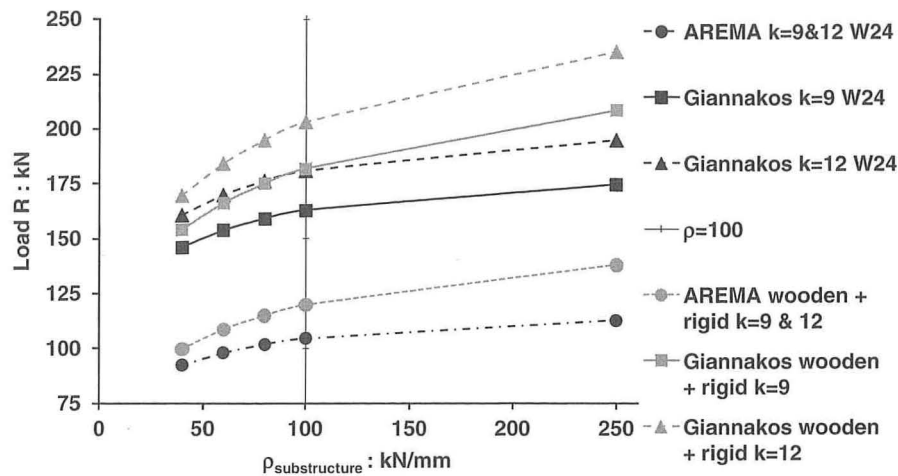


FIGURE 4 Actions on support points of rail ties for wooden ties with rigid fastening (either cut or elastic spike) without pad and concrete tie with resilient fastening W24 (tension clamp Ski-24 and rail pad Zw700WIC) as calculated by AREMA and Giannakos methods.

fastening on wooden ties. For a worse running rail table (e.g.,  $k = 12$ ) the attenuation of the impact loads resulting from the W24 is much greater.

The actions on each tie are higher in the case of the rigid fastening compared with the W24 fastening for both the AREMA and Giannakos (4) methods:

- AREMA: for  $\rho_{\text{subtr}} = 40 \text{ kN/mm}$ , the action on each tie is higher by 8.1%;
- AREMA: for  $\rho_{\text{subtr}} = 250 \text{ kN/mm}$ , the action on each tie is higher by 22.24%;
- Giannakos: for  $\rho_{\text{subtr}} = 40 \text{ kN/mm}$ , the action on each tie is higher by 5.65%; and
- Giannakos: for  $\rho_{\text{subtr}} = 250 \text{ kN/mm}$ , the action on each tie is higher by 20.71%.

The stresses on the ballast at the seating surface of each tie (Figures 5, 6, and 7) are also higher for the rigid fastening compared with the W24 fastening for both methods:

- AREMA: for  $\rho_{\text{subtr}} = 40 \text{ kN/mm}$ , the action on each tie is higher by 8.09%;
- AREMA: for  $\rho_{\text{subtr}} = 250 \text{ kN/mm}$ , the action on each tie is higher by 22.24%;
- Giannakos: for  $\rho_{\text{subtr}} = 40 \text{ kN/mm}$ , the action on each tie is higher by 8.54%; and
- Giannakos: for  $\rho_{\text{subtr}} = 250 \text{ kN/mm}$ , the action on each tie is higher by 25.50%.

The increase of the stresses on the track ballast bed leads to a significant increase of the annual maintenance cost according to the

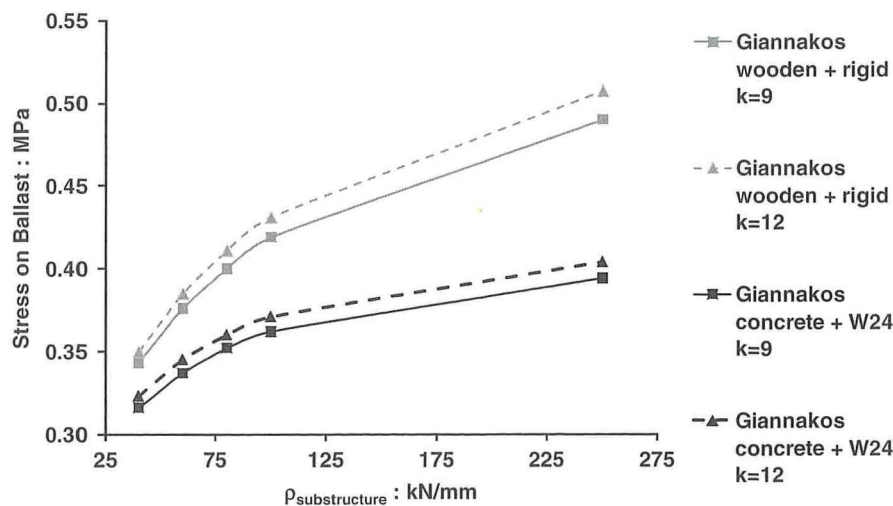


FIGURE 5 Stress on ballast according to Giannakos method for wooden tie with rigid fastening (either cut or elastic spike) without pad and concrete tie with resilient fastening W24 and rail pad Zw700WIC.

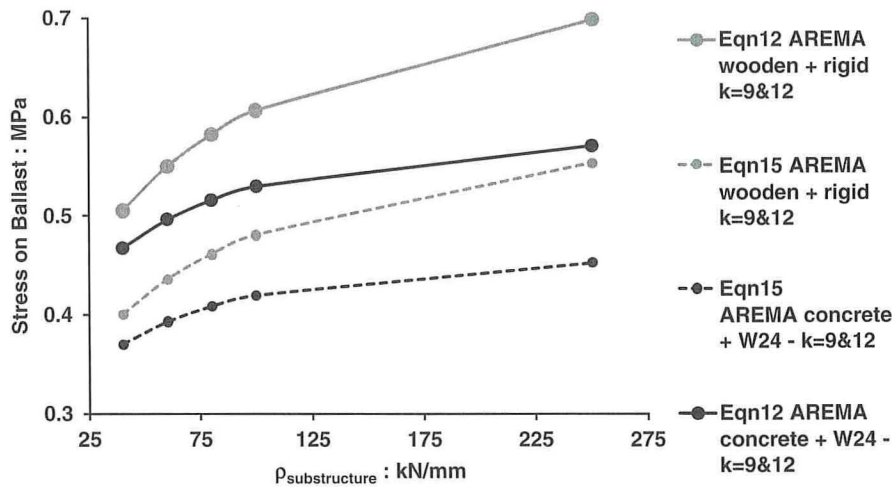


FIGURE 6 Stress on ballast according to AREMA method for wooden tie with rigid fastening (either cut or elastic spike) without pad and concrete tie with resilient fastening W24 and rail pad Zw700WIC (Eqn = equation).

AASHTO road test (applying the fourth-power rule), from  $1.08^4 = 136\%$  to  $1.255^4 = 248\%$ .

Similar favorable results from using very resilient rail pads instead of stiff ones are cited in Harrison and Ahlbeck for impact loading for heavy haul rail (17) and in Eisenmann and Rump for the ballasted track for high-speed rail (i.e., for speeds  $\geq 250$  km/h [155.38 mph] and axle load = 22.5 t [24,806 lb]) (18). For the case of ballasted track for high speed, Eisenmann et al. also recommend increasing the track's elasticity to reduce maintenance cost by lengthening the time intervals for tamping and lining work by a factor of 2.4 (240%) when applying the fourth-power rule, as described above (19). FRA remarks that "a pad equivalent to the 9-mm pad in new condition (500,000 lb/in. = 90 t/in.) is the best choice to prevent cracks under

all but the most extreme conditions" (20). The 9-mm tie pad of the 1980s was much stiffer than today's Zw700WIC pad (with a W24 fastening), which suggests that the stress at the seating surface of the tie (in the case of the W24 fastening) will be much smaller than with the stiffer pads.

The use of resilient fastenings instead of rigid ones leads to an impressive reduction of the annual maintenance cost of the railway track infrastructure. The technological progress in pad production led to a new generation of resilient products of better quality. According to Greek and German regulations the pads should provide a life cycle of 20 years (21). The corresponding very resilient fastening clip should provide a life cycle of 30 years. Moreover, according to research performed by the European Rail Research Institute of the International

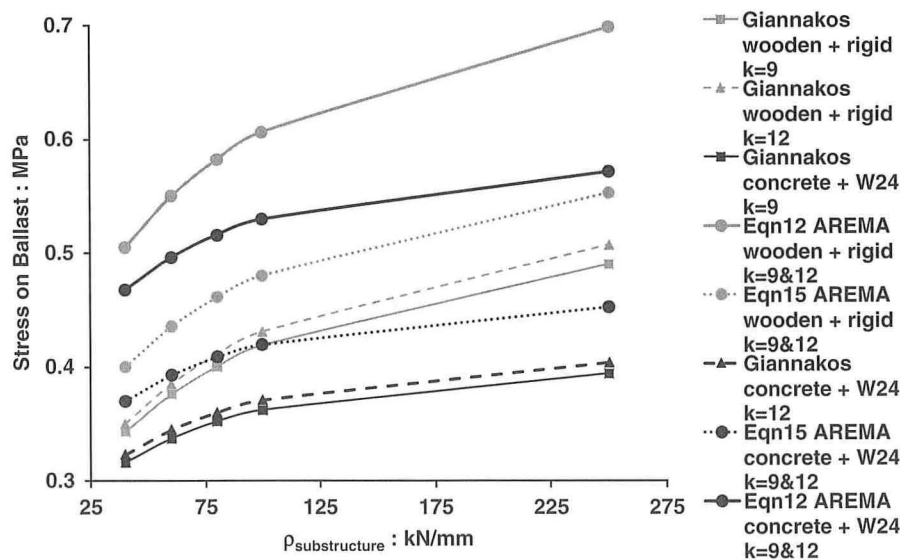


FIGURE 7 Stress on ballast according to Giannakos and AREMA methods for wooden tie with rigid fastening (either cut or elastic spike) without pad and concrete tie with resilient fastening W24 and rail pad Zw700WIC.

Union of Railways, the W clip (e.g., the W14 clip) stops at a deflection of 2 mm rail tilt and presents a very sufficient strength (22, 23). The development of channeled or studded pads permitted the increase of the pad's elasticity (the production of very resilient pads) at the order of 45 to 50 kN/mm (256.96 to 285.51 kips/in.) because of air circulation, which facilitates heat removal and (probable) water sewerage (4). This should imply an avoidance of the pad's degradation or blowing out and rail-seat abrasion. In Europe (with high values of the dynamic component of the acting load) no rail-seat abrasion has been reported with the use of very resilient rail pads.

A comparison between the loads on track imposed by heavy haul traffic with slow speeds (wheel load 17.69 t [39,006 lb; 35.38 t per axle]; maximum speed 60 mph [96.6 km/h]) and high-speed traffic with normal axle loads (250 km/h [155.38 mph] and 22.5 t [49.60 kip]) yields results more adverse for the case of high-speed lines. The results for the heavy haul are comparable to loading for  $V = 140$  km/h (87.01 mph) and 22.5 t (49.60 kip) (12), mainly as a result of the dynamic component of the acting load (see the term in the square root in Equation 16) being directly proportional to the maximum operational speed. Ballast churning and degradation is independent of the tie's material but dependent on the track's total stiffness (6). The exerted pressure on the ballast bed at the tie's seating surface, which is mainly dependent on the total track stiffness coefficient, is the design criterion for a railway track, either in normal or heavy haul circulation. Finally, wooden ties, besides the increased initial cost, present an increased annual maintenance cost of 25% to 40% (24) and a reduced life cycle of 60% to 80% in comparison to concrete ties.

## CONCLUSION

Compared with rigid fastenings, modern high-resilience fastenings significantly reduce the actions on the concrete ties and track superstructure as well as the stress on ballast, and their use must be obligatory on modern railway tracks since they eliminate the problems created by the poor geotechnical conditions of the track substructure. The application of the proposed method for heavy haul rail transportation in the United States with axle loads of 35.38 t (39,006 lb) on a track equipped with W24 fastenings results in stress on the ballast bed at the seating surface of the ties smaller than the relevant stresses if rigid fastenings were used. The W24 fastening gives very good results for heavy haul rail transport. Increases of the stresses on the ballast bed of the track from 8% to 25.5% significantly affect annual maintenance costs. When the fourth-power rule of the AASHTO road test [decrease in track geometry quality = (increase in stress on ballast bed)<sup>4</sup>] is applied, maintenance costs could increase from 136% to 248% depending on the quality of the substructure, from pebbly subgrade to tunnels with a rocky bottom and a small height of the ballast bed cross section.

## REFERENCES

1. Winkler, E. *Die Lehre von der Elastizität und Festigkeit (The Theory of Elasticity and Stiffness)*. H. Dominicus, Prague, 1867.
2. Weaver, W., Jr., S. P. Timoshenko, and D. H. Young. *Vibration Problems in Engineering*, 5th ed. John Wiley & Sons, Hoboken, N.J., 1990.
3. Zimmermann, H. *Die Berechnung des Eisenbahnüberbaues*. Verlag von Wilhelm Ernst & Sohn, Berlin, 1941.
4. Giannakos, K. *Actions on the Railway Track*. Papazissis Publications, Athens, Greece, 2004. [www.papazisi.gr](http://www.papazisi.gr).
5. Esveld, C. *Modern Railway Track*. MRT Productions, Zaltbommel, Netherlands, 2001.
6. Giannakos, K. Loads on Track, Ballast Fouling and Life Cycle Under Dynamic Loading in Railways. *Journal of Transportation Engineering*, Vol. 136, No. 12, 2010, pp. 1075–1084.
7. Giannakos, K., and A. Loizos. Evaluation of Actions on Concrete Sleepers as Design Loads: Influence of Fastenings. *International Journal of Pavement Engineering*, Vol. 11, No. 3, 2010, pp. 197–213.
8. Eisenmann, J., and L. Mattner. Auswirkung der Oberbaukonstruktion auf die Schotter und Untergrundsbeanspruchung. *Eisenbahningenieur*, Vol. 35, No. 3, 1984, pp. 99–107.
9. Eisenmann, J. Verformungsverhalten der Schiene. Auswirkungen auf die Oberbaubeanspruchung, Lagestabilität und Störungsfunktion. *ZEV-Glasers Annalen*, Berlin, 1980, pp. 37–41.
10. Eisenmann, J. *Schotteroberbau: Möglichkeiten und Perspektiven für die Moderne Bahn*. Giesel-Verlag, Isernhagen, Germany, 1988.
11. *Manual for Railway Engineering*. Revised version. American Railway Engineering and Maintenance-of-Way Association, Lanham, Md., 2005.
12. Giannakos, K. The Use of Strain-Attenuating Tie Pads and Its Influence on the Rail Seat Load in Heavy-Haul Railroads. Presented at 2010 Joint Rail Conference, Urbana-Champaign, Ill., 2010.
13. *Precast Concrete Railway Track Systems*. Special edition. Task Group 6.5 of Committee 6 for Prefabrication, Federation Internationale du Beton, Lausanne, Switzerland, 2006.
14. Ahlbeck, D. R., H. C. Meacham, and R. H. Prause. The Development of Analytical Models for Railroad Track Dynamics. In *Railroad Track Mechanics and Technology: Proceedings of a Symposium Held at Princeton University, April 21–23, 1975* (A. D. Kerr, ed.), Oxford, United Kingdom; New York, Pergamon Press, 1978, pp. 239–264.
15. Hay, W. *Railroad Engineering*. John Wiley & Sons, Hoboken, N.J., 1982.
16. *Concrete Railway Sleepers*. Federation Internationale Precontrainte, London, 1984.
17. Harrison, H. D., and D. R. Ahlbeck. Railroad Track Structure Performance Under Wheel Impact Loading. In *Transportation Research Record 1131*, TRB, National Research Council, Washington, D.C., 1987, pp. 81–88.
18. Eisenmann, J., and R. Rump. Ein Schotteroberbau für hohe Geschwindigkeiten. *ETR*, Vol. 46, No. 3, 1997, pp. 98–108.
19. Eisenmann, J., G. Leykauf, and L. Mattner. Vorschläge zur Erhöhung der Oberbauelastizität. *ETR*, Vol. 43, No. 7–8, 1994, pp. 491–494.
20. Office of Research and Development. *Investigation of the Effect of Tie Pad Stiffness on the Impact Loading of Concrete Ties in the Northeast Corridor*. Final report. FRA, U.S. Department of Transportation, 1983.
21. *Elastische Zwischenlagen*. DB-TL 918 235. Munich, Germany, 1994.
22. *Calculation of Twist and Displacements in UIC60 Rail Using Rail Fastening System W and Rail Pads Zw687a and Zw700*. D 170/DT 282. European Rail Research Institute, Utrecht, Netherlands, 1994.
23. *Stresses on the Fastening System Under the Action of Wheel Loads*. D 170/DT 302. European Rail Research Institute, Utrecht, Netherlands, 1994.
24. *Facteurs agissant sur le coût de l'entretien de la voie et leur importance relative* (3rd ed.). International Union of Railways, Paris, 1992.

*The Railway Maintenance Committee peer-reviewed this paper.*

# Biobjective Optimization Model for Maintenance and Renewal Decisions Related to Rail Track Geometry

A. Ramos Andrade and P. Fonseca Teixeira

**This study developed a biobjective optimization model for planning maintenance and renewal actions related to track geometry in a railway network. The problem was modeled as a biobjective integer optimization problem from the perspective of the infrastructure manager. Two objective functions were minimized: (a) the total costs of planned maintenance and renewal actions and (b) the total number of train delays caused by speed restrictions. A small example for a simple network was analyzed in which the optimal Pareto frontier was found through a simulated annealing technique.**

The ever-increasing demand to obtain value for money in railway infrastructure investments has contributed to more conscious and transparent decision making in maintenance processes. In fact, the European railway sector—a recent vertical separated sector—demands more cost-effective tools to enhance decision-making processes related to maintenance and renewal actions. Considering that maintenance costs for a rail track subsystem may represent 55% of total maintenance costs for a high-speed line (1) and that several best-practice studies conducted by the Office of Rail Regulation consisting of international visits to infrastructure managers reported an expected decrease of existing maintenance costs on the order of 20% to 30% through the development of life-cycle cost tools for infrastructure maintenance (2), more research concerning optimization of maintenance and renewal actions is needed. Moreover, decision support systems in railway infrastructure may help to reduce maintenance and renewal costs while mitigating the impacts of train delays and allowing more deliberate decision making in maintenance and renewal actions.

The outline of this paper is as follows: the first section introduces the need for optimizing rail track maintenance and renewal actions related to track geometry, and the second section briefly reviews the state of the art in this domain. The third section explores the model formulation, explaining the two objective functions and identifying the decision variables and model restrictions. The fourth and fifth sections provide, respectively, details about the track geometry degradation phenomenon and the required quality levels. The sixth section provides further details on the assessment of train delays,

and the seventh focuses on the simulated annealing technique. The eighth section explores a network example to run the model, and at last the ninth section highlights the main conclusions and suggests further research.

## REVIEW OF STATE OF THE ART

Budai et al. have presented major contributions to the preventive maintenance scheduling problem for railway infrastructure (3, 4). Their more recent work solved the preventive maintenance scheduling problem by using genetic and memetic algorithms and a two-phase heuristic based on opportunities (4).

Earlier studies addressed the track maintenance scheduling problem as an integer programming problem whose objective was to minimize a weighted combination of expected interference delays and prioritize finishing times (5). Other more straightforward approaches were developed by Simson et al., who proposed an assessment model for rail track maintenance so that track engineers could compare a given maintenance strategy with a base-case scenario without having to fully engage the optimization algorithm (6).

Several other works have optimized different aspects of rail track infrastructure management. Zhao et al. developed separate models to assess the economic life cycle of rail components (7) and to optimize the maintenance actions regarding track geometric quality related to ballast (8). Podofilini et al. put forward a methodology to optimize inspection activities related to rail components (9).

Other sophisticated optimization models have been developed in other areas of railway infrastructure management, especially train operation. Higgins et al. offered a model to optimize rail train scheduling on a single-line track by minimizing a unique objective function combining train delays and operating costs (10). Zhou and Zhong proposed a bicriteria train scheduling model for a high-speed passenger train planning application, minimizing the variation of interdeparture times for high-speed trains and total travel time (11). D'Ariano et al. solved the train scheduling problem in real-time traffic management as a job shop scheduling problem with no-store constraints by using a branch-and-bound algorithm for a bottleneck area of the Dutch rail network (12). Ghoseiri et al. developed a multiobjective passenger train scheduling model minimizing fuel consumption cost and total passenger time (13).

To the best of the authors' knowledge no work has been published in optimizing rail track maintenance and renewal actions by using multiobjective optimization techniques. In response, this paper addresses the planned and preventive maintenance scheduling problem related to rail track geometry by using a biobjective integer formulation, balancing maintenance and renewal costs with train delays.

Center for Urban and Regional Systems, Department of Civil Engineering, Instituto Superior Técnico, Technical University of Lisbon, Av. Rovisco Pais, 1049-001 Lisbon, Portugal. Corresponding author: A. R. Andrade, andrade@civil.ist.utl.pt.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 163–170.  
DOI: 10.3141/2261-19

## MODEL FORMULATION

This section addresses the model formulation. First, details on the notation used are given, followed by a presentation of the two objective functions, the decision variables involved, and the associated restrictions. Proper explanation is included in each subsection to aid the reader's understanding of each part's role in the overall model.

### Notation

The following notation is used:

- $0_1$  = total maintenance and renewal costs (€),
- $0_2$  = total train delay (h),
- $C_{\text{renewal}}$  = renewal cost (€/m),
- $C_{\text{tamping}}$  = tamping cost for a 200-m-long track section (€),
- $\sigma_{ijkl}$  = standard deviation of longitudinal leveling defects (mm) of track section  $k$  from track segment linking node  $i$  to node  $j$  in trimester  $l$ ,
- $\alpha_{ijk}$  = initial standard deviation of longitudinal leveling defects (mm) of track section  $k$  from track segment linking node  $i$  to node  $j$ ,
- $\beta_{ijk}$  = deterioration rate of the standard deviation of longitudinal leveling defects (mm/100 million gross tons [MGT]) of track section  $k$  from track segment linking node  $i$  to node  $j$ ,
- $\gamma_{ijkl}$  = tamping recuperation of the standard deviation of longitudinal leveling defects (mm) of track section  $k$  from track segment linking node  $i$  to node  $j$  in trimester  $l$ ,
- $L_{\text{QN3}}$  = limit for the standard deviation of longitudinal leveling defects for the QN3 quality level (mm),
- $r$  = discount rate,
- $\max_{\text{tamping}}$  = maximum tamping actions possible to conduct in a trimester,
- $\max_{\text{renewal}}$  = maximum renewal actions possible to conduct in a trimester,
- $T_{ijl}$  = tonnage (100 MGT) traversing track segment linking node  $i$  to node  $j$  in trimester  $l$ ,
- $D_{ijl}$  = delays (h) incurred in trimester  $l$  by all trains that traverse track segment linking node  $i$  to node  $j$ ,
- $s$  = speed (km/h),
- $s_{ij}^{\max}$  = maximum speed (km/h) any train can traverse track segment linking node  $i$  to node  $j$ ,
- $s_{ijk}^{\max}$  = maximum speed (km/h) any train can traverse section  $k$  from track segment linking node  $i$  to node  $j$ ,
- $s_{ij}^t$  = maximum speed (km/h) a given train  $t$  can traverse track segment linking node  $i$  to node  $j$ ,
- $s^t$  = maximum speed (km/h) of train  $t$ ,
- $s_{ijl}$  = maximum speed (km/h) any train can traverse track segment linking node  $i$  to node  $j$  in trimester  $l$ ,
- $L_{ij}$  = total length (m) of track segment linking node  $i$  to node  $j$ , and
- $X_{ijkl}, Y_{ijl}, Z_{ijl}$  = decision variables.

### Formulation of Objective Functions

This subsection addresses the formulation of the objective functions. The resulting formulation is a biobjective nonlinear integer programming model. Two main objectives are defined: the first minimizes maintenance (tamping actions) and renewal costs; the second minimizes train delays. These two objectives conflict, for if maintenance

and renewal actions are not performed to ensure safety and passenger comfort, then train speed restrictions would have to be imposed, permissible train speeds would be decreased, and consequently train delays would occur.

The maintenance and renewal cost objective function is expressed as follows:

$$\min 0_1 = \sum_{i,j,l} \frac{C_{\text{renewal}} L_{ij} Y_{ijl} + \sum_k C_{\text{tamping}} X_{ijkl}}{(1+r)^{l/4}} \quad (1)$$

The first objective function is primarily from the perspective of the infrastructure manager. As shown in Equation 1, the infrastructure manager will aim to minimize the total maintenance and renewal costs discounted at a given discount rate  $r$  depending on the trimester  $l$  in which costs are generated.

The planned delays objective function is expressed as follows:

$$\min 0_2 = \sum_{i,j,l} D_{ijl} \quad (2)$$

The second objective function reveals the passenger perspective or a performance perspective (regulator entity) in terms of total delays. The passengers' aim is to see train delays reduced. Details on how each term  $D_{ijl}$  should be computed are provided in the section on assessing train delays.

### Decision Variables

This subsection presents the decision variables ( $X_{ijkl}$ ,  $Y_{ijl}$ ,  $Z_{ijl}$ ) included in the model formulation. Figure 1 shows a simple diagram identifying track sections.

The values for binary variables are calculated as follows:

$X_{ijkl} = 1$  if tamping actions are conducted in section  $k$  from the track segment linking node  $i$  to node  $j$  in trimester  $l$ ;  $X_{ijkl} = 0$  otherwise; and

$Y_{ijl} = 1$  if renewal actions are conducted in all track sections from the track segment linking node  $i$  to node  $j$  in trimester  $l$ ;  $Y_{ijl} = 0$  otherwise.

The integer variable  $Z_{ijl}$  defines the track speed class to which a given track segment linking node  $i$  to node  $j$  in trimester  $l$  belongs, as follows:

$$Z_{ijl} = \begin{cases} 0 & s \leq 80 \text{ km/h} \\ 1 & 80 < s \leq 120 \text{ km/h} \\ 2 & 120 < s \leq 160 \text{ km/h} \\ 3 & 160 < s \leq 200 \text{ km/h} \\ 4 & 200 < s \leq 300 \text{ km/h} \end{cases} \quad (3)$$

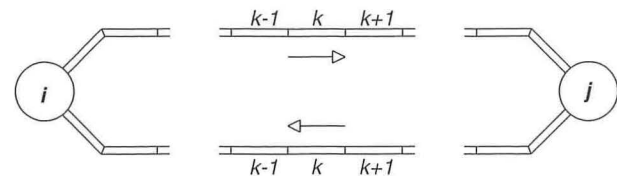


FIGURE 1 Diagram showing track sections from track segments linking (top) node  $i$  to node  $j$  and (bottom) node  $j$  to node  $i$ .

## Restrictions

This subsection presents and discusses the restrictions that a feasible solution should verify:

$$\forall l \quad \sum_{i,j,k} X_{ijk} \leq \max_{\text{tamping}} \quad (4)$$

$$\forall l \quad \sum_{i,j} L_{ij} Y_{ijl} \leq \max_{\text{renewal}} \quad (5)$$

$$\forall i, j, k, l \quad \sigma_{ijk} = \alpha_{ijk} + \beta_{ijk} \sum_{x=1}^l T_{ijx} \left( 1 - \sum_{u=x}^l Y_{iju} \right) - \sum_{x=1}^{l-1} \gamma_{ijkx} X_{ijkx} \left( 1 - \sum_{u=x}^{l-1} Y_{iju} \right) \quad (6)$$

$$\forall i, j, k, l \quad \sigma_{ijk} (1 - X_{ijk} - Y_{ijl}) \leq L_{QN3} \quad (7)$$

$$L_{QN3} = \begin{cases} 1.7 & Z_{ijl} = 4 \\ 2.0 & Z_{ijl} = 3 \\ 2.2 & Z_{ijl} = 2 \\ 2.7 & Z_{ijl} = 1 \\ 3.1 & Z_{ijl} = 0 \end{cases} \quad (8)$$

The first two restrictions (Equations 4 and 5) impose a maximum value of tamping actions in a given trimester and a maximum length of track segment renewals in a given trimester, respectively. These restrictions mimic human, machinery, and budget restrictions. A more detailed analysis of the relationship between these two restrictions, including production rates depending on the tamping machine used, is left for future research.

The next restriction (an equality) in Equation 6 involves the computation of the standard deviation of longitudinal leveling defects for a given track section  $k$  from a track segment linking node  $i$  to node  $j$  in trimester  $l$ , given the associated track geometry degradation parameters for that particular section ( $\alpha_{ijk}$ ,  $\beta_{ijk}$ ,  $\gamma_{ijk}$ ), the accumulated tonnage traversing that track segment since its last renewal ( $T_{ijl}$ ), the tamping decisions related to that section ( $X_{ijk}$ ), and renewal decisions related to that track segment ( $Y_{ijl}$ ).

The model underlying Equation 6 is the linear evolution of the standard deviation of longitudinal leveling defects with accumulated tonnage; it has been widely used as proved by Esvelde (14). To understand Equation 6, it is helpful to suppose that no renewal action has been made for a given track segment linking node  $i$  to node  $j$  for the current trimester  $l$  so that all binary variables  $Y_{ijl}$  are equal to 0. In that particular case the sum of all  $Y_{ijl}$  up to the current trimester is equal to 0, and Equation 6 can be simplified into the following expression:

$$\sigma_{ijk} = \alpha_{ijk} + \beta_{ijk} \sum_{x=1}^l T_{ijx} - \sum_{x=1}^{l-1} \gamma_{ijkx} X_{ijkx} \quad (9)$$

Equation 9 shows that if no renewal action has occurred, the current standard deviation of longitudinal leveling defects in the current trimester  $l$  can be computed as the sum of three terms: the initial standard deviation ( $\alpha_{ijk}$ ), the product between the deterioration rate ( $\beta_{ijk}$ ) and the accumulated tonnage since the last renewal, and a recu-

peration effect caused by tamping actions  $\gamma_{ijk}$ . The section on track geometry degradation provides details on the evidence found in the literature concerning the linear degradation model expressed in Equation 6.

The final restrictions (Equations 7 and 8) should be regarded simultaneously. In fact, they intend to guarantee that the current (trimester  $l$ ) standard deviation of longitudinal leveling defects ( $\sigma_{ijk}$ ) is not above the required limit that results from a necessary consideration of safety and passenger comfort ( $L_{QN3}$ ) depending on the speed restriction imposed ( $Z_{ijl}$ ) for that track segment in the same trimester. The section discussing required geometric quality levels explores and provides more information related to Equation 8. Note that if tamping actions ( $X_{ijk}$ ) or renewal actions ( $Y_{ijl}$ ) are scheduled for the trimester and track section in which the standard deviation of longitudinal defects is assessed ( $\sigma_{ijk}$ ), the final restrictions expressed in Equations 7 and 8 are always verified.

## DEGRADATION OF TRACK GEOMETRY

Degradation of track geometry is usually quantified by five track defects: longitudinal leveling defects, horizontal alignment defects, cant defects, gage deviations, and track twist. Many infrastructure managers tend to sum up all these defects into a track quality index that is typically a function of the standard deviations of each defect and permissible train speed (8, 15). However, the standard deviation for the short wavelength (3 to 25 m) of longitudinal leveling defects is still the crucial parameter for maintenance decisions. This has been confirmed by a recent guide on best practices for optimum track geometry durability (16). Longitudinal leveling defects are defined as the geometrical error in the vertical plane, measured in millimeters from the rail top on the running surface to the ideal mean line of the longitudinal profile.

Many experimental studies have validated a linear relationship between the standard deviation of longitudinal leveling defects and accumulated tonnage. Accumulated tonnage is the sum of all axle loads (in metric tons) of all trains that have run in a given track section since renewal or last tamping action, usually quantified in million gross tons. Using accumulated tonnage instead of time is a more convenient way to assess the evolution of the standard deviation of longitudinal leveling defects as it allows heavily used track sections to be differentiated from track sections that receive less use. Although accumulated tonnage has also been computed by separating passenger from freight train sets, such formulations have been heavily criticized (17), and the method using the sum of all axle loads is used here.

For optimum use of information and control of maintenance and renewal processes, track inspection records should be condensed, usually to records that cover 200-m-long track sections (18). The evolution of the standard deviation of longitudinal leveling defects for each 200-m section can be estimated using the following linear relationship:

$$\sigma_{LD} = c_1 + c_0 \cdot T \quad (10)$$

where

- $\sigma_{LD}$  = standard deviation of longitudinal leveling defects (mm),
- $c_1$  = initial standard deviation measured after renewal or tamping operations (mm),
- $c_0$  = deterioration rate (mm/100 MGT), and
- $T$  = accumulated tonnage since renewal or tamping action (100 MGT).

Although other models may capture the nonlinear characteristics of track quality deterioration (19–21), the linear function is widely used and is employed for the degradation model in the previous section.

Equation 10 is the simplest form of the linear degradation model expressed in Equations 6 and 9, where  $c_1$  and  $c_0$  correspond, respectively, to  $\alpha_{ijk}$  and  $\beta_{ijk}$ . However, Equations 6 and 9 include another parameter ( $\gamma_{ijkl}$ ) related to tamping recuperation. Only a few references were found that cover the improvement caused by tamping actions (21, 22).

Track geometry degradation parameters ( $\alpha_{ijk}$ ,  $\beta_{ijk}$ ,  $\gamma_{ijkl}$ ) describe the evolution of the standard deviation of longitudinal leveling defects. Research conducted for the Portuguese infrastructure manager (REFER) based on an analysis of a sample from the Lisbon–Oporto line showed that it can be assumed that the first two parameters ( $\alpha_{ijk}$ ,  $\beta_{ijk}$ ) follow a correlated bivariate lognormal distribution (23). The same study reported statistically significant spatial correlations between both parameters relative to consecutive track sections in the same direction [ $(\alpha_{ijk}, \beta_{ijk})$  and  $(\alpha_{ijk+1}, \beta_{ijk+1})$ ] and track sections aligned but in opposite directions [ $(\alpha_{ijk}, \beta_{ijk})$  and  $(\alpha_{jik}, \beta_{jik})$ ] (Figure 1). In fact, with regard to the first two parameters, the information cited above is enough to define a random field related to track geometry degradation for planned maintenance purposes. The same reference provides information on the type of track structure in the Lisbon–Oporto case study (23). However, the present optimization model is not restricted to a particular type of track structure and can be adapted for different track structures through the estimation of different track geometry degradation parameters.

Little information is available in the literature regarding the other degradation parameter ( $\gamma_{ijkl}$ ), which refers to the recuperation achieved in the standard deviation of longitudinal leveling defects through tamping actions. Preliminary data from the ongoing research project for the Portuguese infrastructure manager REFER have shown that  $\gamma_{ijkl}$  has a linear relationship with the standard deviation of longitudinal leveling defects before tamping actions occur. This relationship can be expressed as

$$\gamma_{ijkl} = 0.60 \times \sigma_{ijkl} - 0.15 \quad (11)$$

Because Equation 11 is a result of as yet preliminary research, the expression coefficients should be regarded as rough approximations of the real ones. Further research should provide more insight on whether the stated relationship in Equation 11 is appropriate. For instance, these coefficients may have some dependence on  $l$  caused by physical reasons, such as the ballast fatigue phenomenon.

## REQUIRED LEVELS OF GEOMETRIC QUALITY

For safety and passenger comfort reasons, the standard deviation of longitudinal leveling defects should not be higher than a certain limit depending on the train speed. Standard 518 of the International Union of Railways (UIC) (24) sets three geometric track quality levels:

**QN1.** Refers to the value that necessitates monitoring or taking maintenance actions as part of regularly planned maintenance operations,

**QN2.** Refers to the value that requires short-term maintenance action, and

**QN3.** Refers to the value above which it is no longer desirable to maintain the scheduled traffic speed.

**TABLE 1** Standard Deviation Limits of Longitudinal Leveling Defects for Different Train Speeds and Quality Levels QN1 and QN2 According to UIC Standard 518 (24)

Standard Deviation Limits of Longitudinal Leveling Defects	QN1 (mm)	QN2 (mm)
$s \leq 80$ km/h	2.3	2.6
$80 < s \leq 120$ km/h	1.8	2.1
$120 < s \leq 160$ km/h	1.4	1.7
$160 < s \leq 200$ km/h	1.2	1.5
$200 < s \leq 300$ km/h	1.0	1.3

European Standard EN 13848-5 also sets three quality levels that are similar to QN1, QN2, and QN3 but with different names: alert limit, intervention limit, and immediate action limit, respectively (25). Table 1 and Table 2 provide the required limits for different train speeds according to UIC Standard 518 (24) and European Standard EN 13848-5 (25), respectively. As the tables confirm, the limits of train speeds are a little different in UIC Standard 518 for classes in higher speeds, and EN 13848-5 recommends intervals only for the alert limit designation. Although maximum values for defects are limited in both standards, the scheduling of planned maintenance actions are made only on the basis of the standard deviation of longitudinal leveling defects.

The limit values for quality level QN1 are exactly equal to the inferior bound of the recommended interval for the alert limit, although the two classes with higher train speeds are a little different (230 km/h splits them instead of 200 km/h). This means UIC Standard 518 is more demanding than standard EN 13858-5 in terms of the standard deviation of longitudinal leveling defects. In this paper, the limit values given by UIC Standard 518 are assumed.

As an example, for train speeds between 200 and 300 km/h, the QN1 quality level limit for the standard deviation of longitudinal leveling defects is 1.0 mm, and the QN2 quality level limit is 1.3 mm. Table 1 shows how these limits vary depending on train speed. The QN3 quality level is usually set at 130% of the QN2 level (Table 3). It is assumed that the integer decision variable  $Z_{ijl}$  can assume five values depending on the scheduled traffic speed, as Table 3 illustrates.

**TABLE 2** Standard Deviation Limits of Longitudinal Leveling Defects for Different Train Speeds and Alert Limit Quality Level According to European Standard EN 13848-5 (25)

Standard Deviation Limits of Longitudinal Leveling Defects	Alert Limit (mm)
$s \leq 80$ km/h	2.3–3.0
$80 < s \leq 120$ km/h	1.8–2.7
$120 < s \leq 160$ km/h	1.4–2.4
$160 < s \leq 230$ km/h	1.2–1.9
$230 < s \leq 300$ km/h	1.0–1.5

**TABLE 3** Decision Variable ( $Z_{ijl}$ ) and Standard Deviation Limit ( $L_{QNS}$ ) Values for Different Speed Limits

$Z_{ijl}$	Standard Deviation Limits of Longitudinal Leveling Defects	$L_{QNS}$ (mm)
0	$s \leq 80$ km/h	3.1
1	$80 < s \leq 120$ km/h	2.7
2	$120 < s \leq 160$ km/h	2.2
3	$160 < s \leq 200$ km/h	2.0
4	$200 < s \leq 300$ km/h	1.7

## ASSESSING TRAIN DELAYS

This section provides more insight into the intricacies of the computation of train delays. The second objective function to minimize is the total planned train delay (Equation 2).

The main issue is how to compute the delays incurred in trimester  $l$  by all trains that traverse the track segment linking node  $i$  to node  $j$  ( $D_{ijl}$ ). To compute  $D_{ijl}$  as a function of  $Z_{ijl}$  some simplifications are made. Each track segment linking node  $i$  to node  $j$  has a maximum permissible speed ( $s^{\max}$ ) depending on the features of its infrastructure, which should be the minimum speed of all track sections composing that track segment:

$$s_{ij}^{\max} = \min_k s_{ijk}^{\max} \quad (12)$$

Each train  $t$  has a maximum speed ( $s^t$ ) regardless of the features of the infrastructure. The actual maximum speed that a given train  $t$  can traverse the track segment linking node  $i$  to node  $j$  is given by the following expression:

$$s_{ij}^t = \min\{s_{ij}^{\max}, s^t\} \quad (13)$$

Nevertheless, speed restrictions may have been imposed in a given trimester  $l$  as a result of an undesirable increase of the standard deviation of longitudinal leveling defects, and thus the maximum permissible speed of each train  $t$  traversing the track segment linking node  $i$  to node  $j$  in a given trimester  $l$  is given by

$$s_{ijl}^t = \min\{s_{ij}^t, s_{ijl}(Z_{ijl})\} \quad (14)$$

Values for  $s_{ijl}(Z_{ijl})$  are determined as follows:

$$s_{ijl}(Z_{ijl}) = \begin{cases} 300 \text{ km/h} & Z_{ijl} = 4 \\ 200 \text{ km/h} & Z_{ijl} = 3 \\ 160 \text{ km/h} & Z_{ijl} = 2 \\ 120 \text{ km/h} & Z_{ijl} = 1 \\ 80 \text{ km/h} & Z_{ijl} = 0 \end{cases} \quad (15)$$

As a simplification, it is assumed that acceleration and breaking times are negligible, and delays caused by speed restrictions for the track segment linking node  $i$  to node  $j$  in a given trimester  $l$  are computed through the following expression:

$$D_{ijl} = \sum_t L_{ij} \left( \frac{1}{s_{ijl}^t} - \frac{1}{s_{ijl}^{\max}} \right) \quad (16)$$

where  $L_{ij}$  is the length of the track segment linking node  $i$  to node  $j$ .

The second objective function ( $0_2$ ) depends on the decision variable  $Z_{ijl}$  and can be expanded as

$$\begin{aligned} 0_2 &= \sum_{i,j,l} D_{ijl} = \sum_{i,j,l} \sum_t L_{ij} \left( \frac{1}{s_{ijl}^t} - \frac{1}{s_{ijl}^{\max}} \right) \\ &= \sum_{i,j,l} \sum_t L_{ij} \left( \frac{1}{s_{ijl}^t} - \frac{1}{\min\{s_{ijl}^t, s_{ijl}(Z_{ijl})\}} \right) \end{aligned} \quad (17)$$

This assessment of train delays only considers delays caused by planned maintenance and renewal actions and does not consider other types of train delays (e.g., delays caused by unplanned maintenance actions, primary train delays caused by operators' delays, and secondary train delays caused by the propagation of primary train delays in a given network). These aspects are particularly complex to model without using specific simulation software (e.g., Opentrack). The train delays quantified in the present formulation are called planned delays because they are caused by a given planned maintenance and renewal strategy. It remains for further research to overcome these simplifications.

## SIMULATED ANNEALING

This section briefly presents the simulated annealing technique as a metaheuristic for optimization problems. The next section shows how this technique can be used to solve a small example of the proposed model.

Simulated annealing is a widely used metaheuristic based on the Metropolis algorithm. The main idea of this technique is to iteratively and randomly search for a better solution, although in some instances the technique allows solution changes that worsen the current solution in an attempt to reduce the probability of becoming trapped in a locally suboptimal solution.

Kirkpatrick et al. (26) and Cerny (27) first used the simulated annealing technique suggested by Metropolis et al. (28) to solve combinatorial problems of the well-known traveling salesman problem. The name of the technique exploits an analogy between the way in which a metal cools and freezes into a minimum-energy crystalline structure (the annealing process) and the search for a minimum in a more general system. In the annealing process, the temperature is first raised and then gradually decreased to a very low level, ensuring that enough time (iterations) is spent at each temperature level. Although enthusiasm for the technique increased when Geman and Geman proved that simulated annealing converges to the global optimum if annealing is sufficiently slow (29), Bandyopadhyay et al. report that few researchers have attempted to extend the simulated annealing technique to multiobjective optimization problems (30). In fact, most earlier attempts, such as Czyzak and Jaskiewicz, were based on the construction of a single-objective function through the combination of the different objectives into one using a weighted sum approach (31).

In simple terms, domination is the most important concept in multi-objective optimization. In a minimization problem, a given solution  $x$  dominates  $x'$  if  $\forall m f_m(x) \leq f_m(x')$  and  $\exists m f_m(x) < f_m(x')$ . In other words, a given solution  $x$  dominates  $x'$  if for all objectives solution

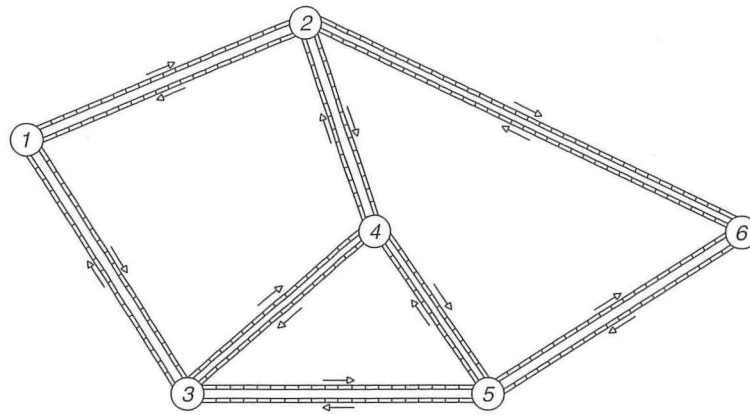


FIGURE 2 Diagram of network example.

$x$  is just as good as solution  $x'$  and for at least one objective function, solution  $x$  is better than  $x'$ .

A multiobjective optimization problem contains a set of non-dominated solutions called the Pareto optimal set. These solutions are considered to be equally important, and only with further information related to the decision maker's preferences can the best solution (adapted to that particular decision maker) be found.

The main general steps of simulated annealing algorithms for multiobjective optimization are as follows:

1. Find a feasible solution ( $x$ ) and set it as the current solution.
2. Set the cooling parameters, including the initial temperature  $T$ , the cooling rate  $r$ , and the epoch length ( $\text{len}$ ).
3. For  $1 \leq i \leq \text{len}$ , take the following steps:
  - Pick a random neighbor ( $x'$ ) that can be a feasible solution;
  - Test if  $x'$  dominates all nondominated solutions  $x$ ;
  - If  $x'$  dominates all nondominated solutions  $x$ , include  $x'$  in the nondominated solutions set; and
  - If  $x'$  does not dominate all nondominated solutions  $x$ , include  $x'$  in the nondominated solutions set with probability  $p(T)$ .
4. Set  $T = r \times T$  and repeat Step 3.
5. Return the nondominated solutions set.

The cooling parameters mentioned in Step 2 are quantities that should be fitted in order to improve the computer efficiency of the simulated annealing technique. The present paper does not consider these aspects in detail and leaves them for further research.

## MODEL APPLICATION IN A SMALL EXAMPLE

A small example was built in order to show how the present model can support infrastructure manager renewal and maintenance decisions. All the following information is needed:

- Track geometry degradation parameters ( $\forall i, j, k, \alpha_{ijk}, \beta_{ijk}$ ),
- Tamping recuperation ( $\gamma_{ijk}$ ),
- Tamping and renewal costs ( $C_{\text{tamping}}, C_{\text{renewal}}$ ) and discount rate ( $r$ ) ( $\text{€}1 = \$1.278$ , July 2010 dollars),
- Tonnage ( $T_{ijl}$ ),
- Track segment length ( $L_{ij}$ ),
- Maximum permissible speed of track segment ( $s_{ij}^{\text{max}}$ ) and maximum train permissible speed ( $s'_{ij}$ ), and

- Maximum of tamping actions ( $\text{max}_{\text{tamping}}$ ) and maximum length of track segment renewals ( $\text{max}_{\text{renewal}}$ ).

Figure 2 illustrates and the following table provides values for some of the above information related to the example:

Variable	Value
$C_{\text{tamping}} (\text{€})$	2,000
$C_{\text{renewal}} (\text{€})$	150,000
$\text{max}_{\text{tamping}}$	30
$\text{max}_{\text{renewal}} (\text{km})$	5
$r (\%)$	30

Values must be assumed for track segment lengths ( $L_{ij}$ ) and the total tonnage ( $T_{ijl}$ ) from all trains that traverse the track segment linking node  $i$  to node  $j$  in trimester  $l$  (Table 4). It is assumed that tonnage is constant regardless of trimester. The right-hand column in Table 4

TABLE 4 Values for Each Track Segment Linking Node  $i$  to Node  $j$  in Example

$i$	$j$	Length (km) ( $L_{ij}$ )	Tonnage per Trimester (MGT) ( $T_{ij}$ )	Number of Trains per Trimester
1	2	4.0	2.5	7,500
2	1	4.0	2.5	7,500
1	3	4.0	3.75	11,250
3	1	4.0	3.75	11,250
2	4	2.8	3	9,000
4	2	2.8	3	9,000
3	4	3.2	2	6,000
4	3	3.2	2	6,000
3	5	4.0	3	9,000
5	3	4.0	3	9,000
4	5	2.4	1.25	3,750
5	4	2.4	1.25	3,750
5	6	4.0	2.5	7,500
6	5	4.0	2.5	7,500
2	6	6.8	3	9,000
6	2	6.8	3	9,000

NOTE: Maximum speed ( $s'_{ij}$ ) for all cases is 220 km/h.

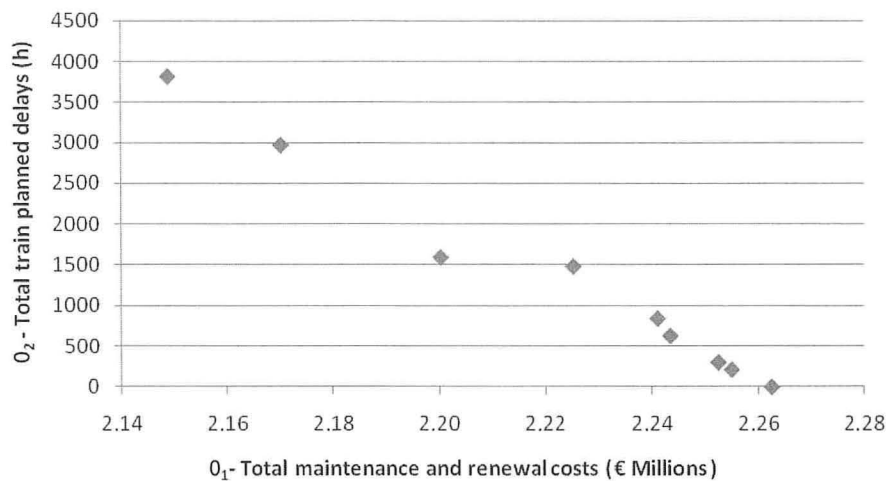


FIGURE 3 Approximation of optimal Pareto solution set.

includes values for the number of trains per trimester for each track segment. For simplification, it is assumed that traffic is homogeneous and that each train has the same maximum speed for each segment.

The track geometry degradation parameters ( $\alpha_{ijk}$ ,  $\beta_{ijk}$ ) can be simulated as a random field for planned maintenance purposes. Values were estimated for each track section of the present network example according to the findings of the uncertainty associated with  $\alpha_{ijk}$ ,  $\beta_{ijk}$  and spatial correlations. The third parameter, tamping recuperation ( $\gamma_{ijk}$ ), was calculated according to Equation 11.

Figure 3 provides the set of nondominated solutions found for a time horizon of 30 years. Because the simulated annealing technique provides only an approximation of the Pareto optimal set, this figure should be regarded as an approximation of it.

Figure 3 shows one nondominated solution with no train delays ( $O_1 = €2.262$  million and  $O_2 = 0$ ), but this solution has the highest cost. Although only nine nondominated solutions were found, the figure gives a rough approximation of the relationship between total costs and associated delays in the optimum set.

Further research is needed to understand how the simulated annealing technique picks a random neighbor given a current solution. In fact, this simulation example showed that in the thousands of feasible solutions simulated, there were only nine nondominated solutions. One possibility for further research would be assessing the computational efficiency of the simulated annealing technique compared with a genetic algorithm.

## CONCLUSIONS AND FURTHER RESEARCH

This paper presents a biobjective model to optimize the planned maintenance and renewal actions related to track geometry. Two main objectives are minimized: total maintenance and renewal costs and planned train delays. The resulting model incorporates a track degradation model and the required safety and comfort limits depending on the train speed. A small example suggests how simulated annealing techniques can be used to find the Pareto optimal set of solutions. More research in the performance of simulated annealing techniques to approximate the Pareto optimal set of nondominated solutions is needed.

Because of the strong influence that the degradation model used for the standard deviation of longitudinal leveling defects has in the

optimization results, the authors believe that sensitivity tests of the assumptions concerning the tamping recuperation and the degradation model itself should be conducted. A study of the variability of the Pareto optimal set of nondominated solutions for different network configurations and different degradation parameters is also desirable.

Finally, a more comprehensive assessment of both objective functions should be pursued in order to include all associated costs for additional maintenance and renewal actions (e.g., rail replacements, grinding actions) and delay impacts in train operation (e.g., planned and unplanned train delays caused by all maintenance and renewal actions).

## ACKNOWLEDGMENTS

The authors are grateful for the support and collaboration of the Portuguese railway infrastructure manager, REFER, EPE, as well as for the support of the Portuguese Foundation for Science and Technology, FCT, and MIT Portugal. Finally, the authors thank the reviewers for their comments and suggestions to improve the final manuscript.

## REFERENCES

1. López-Pita, A., P. F. Teixeira, C. Casas, A. Bachiller, and P. A. Ferreira. Maintenance Costs of High-Speed Lines in Europe. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2043, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 13–19.
2. Office of Rail Regulation. *ORR Best Practice Study. A Report on the Programme of International Visits Carried Out by ORR Between July–October 2007*. Summary report, Version 3. 2008. <http://www.rail-reg.gov.uk>. Accessed June 20, 2010.
3. Budai, G., D. Huisman, and R. Dekker. Scheduling Preventive Maintenance Activities. *Journal of the Operational Research Society*, Vol. 57, 2006, pp. 1035–1044.
4. Budai, G., R. Dekker, and U. Kaymak. *Genetic and Memetic Algorithms for Scheduling Railway Maintenance Activities*. Econometric Institute Report EI 2009-30. Erasmus University, Rotterdam, Netherlands, 2009.
5. Higgins, A., L. Ferreira, and M. Lake. Scheduling Rail Track Maintenance to Minimize Overall Delays. *Proc., 14th International Symposium on Transportation and Traffic Theory*, Jerusalem, Israel, 1999.

6. Simson, S. A., L. Ferreira, and M. H. Murray. Rail Track Maintenance Planning: An Assessment Model. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1713, TRB, National Research Council, Washington, D.C., 2000, pp. 29–35.
7. Zhao, J., A. H. C. Chan, C. Roberts, and A. B. Stirling. Assessing the Economic Life of Rail Using a Stochastic Analysis of Failures. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, Vol. 220, No. 2, 2006, pp. 103–111.
8. Zhao, J., A. H. C. Chan, A. B. Stirling, and K. B. Madelin. Optimizing Policies of Railway Ballast Tamping and Renewal. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1943, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 50–56.
9. Podofilini, L., E. Zio, and I. Vatn. Risk-Informed Optimization of Railway Tracks Inspection and Maintenance Procedures. *Reliability Engineering and System Safety*, Vol. 91, 2006, pp. 20–35.
10. Higgins, A., E. Kozan, and L. Ferreira. Optimal Scheduling of Trains on a Single-Line Track. *Transportation Research Part B*, Vol. 30, No. 2, 1996, pp. 147–161.
11. Zhou, X., and M. Zhong. Bicriteria Train Scheduling for High-Speed Passenger Railroad Applications. *European Journal of Operational Research*, Vol. 167, 2005, pp. 752–771.
12. D'Ariano, A., D. Pacciarelli, and M. Pranzo. A Branch and Bound Algorithm for Scheduling Trains in a Railway Network. *European Journal of Operational Research*, Vol. 183, 2007, pp. 643–657.
13. Ghoseiri, K., F. Szidarovszky, and M. J. Asgharpour. A Multi-Objective Train Scheduling Model and Solution. *Transportation Research Part B*, Vol. 38, 2004, pp. 927–952.
14. Esveld, C. *Modern Railway Track*, 2nd ed. MRT Productions, Zaltbommel, Netherlands, 2001.
15. El-Sibaie, M., and Y.-J. Zhang. Objective Track Quality Indices. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1863, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 81–87.
16. *Best Practice Guide for Optimum Track Geometry Durability*. International Union of Railways, Paris, 2008.
17. *UIC Leaflet 714: Classification of Lines for the Purpose of Track Maintenance*, 4th ed. International Union of Railways, Paris, 2009.
18. Esveld, C. Computer-Aided Maintenance and Renewal of Track. 1990 ASME/IEEE Joint Railroad Conference, 1990, pp. 165–170.
19. Riessberger, K. Extending Maintenance Cycles and Ballast Life. *Proc., 7th International Heavy Haul Conference*, Brisbane, Australia, 2001.
20. Ubalde, L., A. López-Pita, P. Teixeira, A. Bachiller, and I. Gallego. Track Deterioration in High-Speed Railways: Influence of Stochastic Parameters. *8th International Conference on Railway Engineering*, London, 2005, pp. 54–62.
21. Veit, P. Track Quality: Luxury or Necessity? Maintenance and Renewal. *Journal of Railway Technical Review*, Special Issue, July 2007, pp. 8–12.
22. Lichtberger, B. *Track Compendium: Formation, Permanent Way, Maintenance, Economics*. Eurailpress, Hamburg, Germany, 2005.
23. Andrade, A. R., and P. F. Teixeira. Uncertainty in Rail-Track Geometry Degradation: Lisbon–Oporto Line Case Study. *Journal of Transportation Engineering*, Vol. 137, No. 3, 2011, pp. 193–200.
24. *UIC Leaflet 518: Testing and Approval of Railway Vehicles from the Point of View of Their Dynamic Behaviour: Safety, Track Fatigue, Ride Quality*, 3rd ed. International Union of Railways, Paris, 2005.
25. *European Standard EN 13848-5: Railway Applications: Track—Track Geometry Quality, Part 5: Geometric Quality Levels*. European Committee for Standardization, Brussels, Belgium, 2008.
26. Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, Vol. 220, No. 4598, 1983, pp. 671–680.
27. Cerny, V. Thermodynamic Approach to the Travelling Salesman Problem. *Journal of Optimization Theory and Applications*, Vol. 45, No. 1, 1985, pp. 41–51.
28. Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, Vol. 21, No. 6, 1953, pp. 1087–1092.
29. Geman, S., and D. Geman. Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 2, 1984, pp. 721–741.
30. Bandyopadhyay, S., S. Saha, U. Maulik, and K. Deb. A Simulated Annealing-Based Multi-Objective Optimization Algorithm: AMOSA. *IEEE Transactions on Evolutionary Computation*, Vol. 12, No. 3, 2008, pp. 269–283.
31. Czyzak, P., and A. Jaskiewicz. Pareto Simulated Annealing: A Metaheuristic Technique for Multiple-Objective Combinatorial Optimization. *Journal of Multi-Criteria Decision Analysis*, Vol. 7, 1998, pp. 34–47.

---

*The Railway Maintenance Committee peer-reviewed this paper.*

# Verification of Box Test Model and Calibration of Finite Element Model

## Evaluation of Railroad Ballast Performance

Kane C. Bennett, Carlton L. Ho, and Hoang Q. Nguyen

The development of innovative models to evaluate railroad ballast performance can reduce railroad maintenance costs and improve train operation and safety. A significant portion of the maintenance costs associated with track transportation systems can be attributed to ballast degradation (fouling). Understanding the mechanisms and the effects of ballast fouling is becoming increasingly important with the increased demand on track transportation systems and as trains become faster and carry more weight. This paper presents the results and findings of a finite element (FE) model to evaluate and validate a railroad ballast box test apparatus developed at the University of Massachusetts, Amherst. FE modeling of the box test was conducted to provide insight for subsequent laboratory box test experiments. Boundary and scale effects of the box test were evaluated through the FE model, and the stress-strain behavior and distribution were explored. The predicted relationship between the elastic modulus and the percentage of fouling of the ballast by wet clay was found through calibration of the FE model. The FE model showed that the boundary conditions of the box governed the behavior of the model, specifically, that the addition of smooth and flexible walls reduced strength capacity and increased plastic deformation of the ballast. The FE model also showed that by increasing the size of the box by approximately 50%, stress concentrations below the tie, which would have led to a failure mechanism resembling punching shear, were avoided.

Railroad track substructure performance and maintenance are becoming increasingly important in light of continuing advancements in train velocities and load-carrying capacities. With the increased interest in high-speed rail and heavy axle load freight, track conditions will become more constrained (1–3). With maintenance costs estimated at between 26.6 and 80.3 thousand dollars per mile of track, ensuring that the track is safe for operation can be expensive (4). Ballast, typically composed of angular, uniformly graded crushed gravel, is prone to degradation (fouling) over time, predominantly from the breaking off of sharp edges (especially during maintenance tamping) and from the introduction of fines into the pore spaces by wind, water, and subgrade intrusion (5). Ballast fouling and settlement account for a large portion of the maintenance costs associated with track transportation systems (5), and the ability to better model, pre-

dict, and quantify ballast performance with respect to fouling can improve safety and reduce maintenance costs (6).

Modeling of ballast behavior with respect to fouling has been accomplished in the past through the development of laboratory box tests, which are meant to represent the interaction of a short section of railroad tie directly beneath the rail with the supporting ballast substructure (7–10). The box test is a plane strain test on the ballast and tie system (see Figures 1 and 2). Lu and McDowell recognized the importance of modeling the behavior of box tests and did so by using a discrete element method (12). Fouled ballast, however, does not lend itself to the discrete element method as well as clean ballast because of the fine-grained material filling the void spaces. Geotechnical analysis using finite element (FE) software is becoming increasingly common in practice, and hence should be explored. The objective of the present research was to develop an FE model with the PLAXIS (13) FE program in order to simulate these box test results, calibrate the FE model with the box test data, and investigate and validate the box test model with the developed FE model. This research was conducted to assist in the planning of a program of subsequent fouled ballast box tests to be done in conjunction with triaxial shear tests at the University of Massachusetts, Amherst.

### BOX TEST MODEL

The analysis was conducted on data from box tests at the University of Massachusetts (7, 8). The box test model of the University of Massachusetts, Amherst, was first developed in the early 1980s. The first box was made of wood with steel angle iron reinforcement. Its dimensions were relatively small (see Figure 2), and the effect of the imposed rigid boundary was intended to be mitigated by various arrangements and types of flexible rubber linings, often covered by another thin layer of sheet metal or wood. The intention of the first box test model was twofold: (a) to evaluate the development of horizontal stresses in the ballast layer under cyclic loading conditions similar to those imposed by train traffic and (b) to examine the change in behavior of the ballast as it degrades (and hence fouls) under these loads (7). The box was filled with crushed trap rock, a commonly used ballast material, and a simulated stiff section of tie spanned the minor horizontal axis. The box was instrumented with stress and strain gauges and tested under static and cyclic loads between 50 and 4,000 lbs (0.22 and 17.8 kN).

A second box test apparatus was built at the University of Massachusetts, Amherst, in the late 1980s. The new box was made to the same dimensions as the original box, but it was constructed of sheet metal with angle iron welded to the exterior to increase stiffness. A firm rubber lining was installed on the sides and base and covered

---

Department of Civil and Environmental Engineering, University of Massachusetts, Amherst, 28 Marston Hall, 130 Natural Resources Road, Amherst, MA 01003. Corresponding author: K. C. Bennett, kbennett@ecs.umass.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 171–177.  
DOI: 10.3141/2261-20

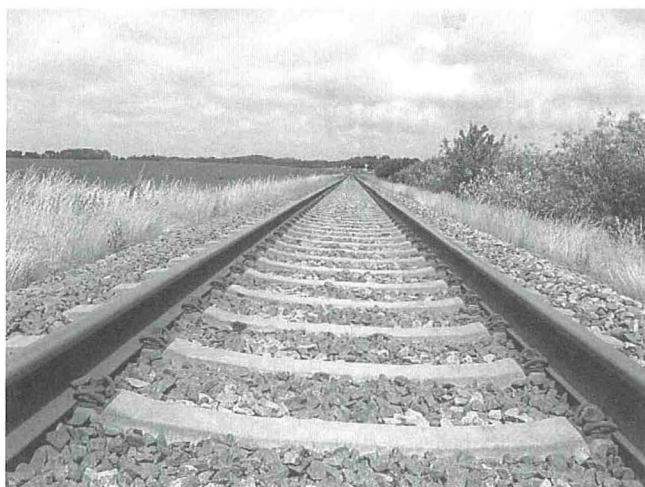


FIGURE 1 Section of railroad track showing rail, tie, and ballast structure.

with sheets of thin stainless steel. The box was built for the purpose of evaluating the effect of fouling on ballast behavior under long-term cyclic loading (8). Dolomite limestone ballast material was tested at various degrees of fouling, from 0% (clean ballast) to 40% (ballast composed of 40% fines passing a No. 200 sieve). For the present study, tests involving wet clay as the fouling material were examined because, of all the fouling materials used, wet clay showed the greatest potential for changing the behavior of the ballast. The effect of the wet clay fouling material on the ballast behavior can be attributed to its low hydraulic conductivity, relatively high cohesive strength, and low angle of internal friction.

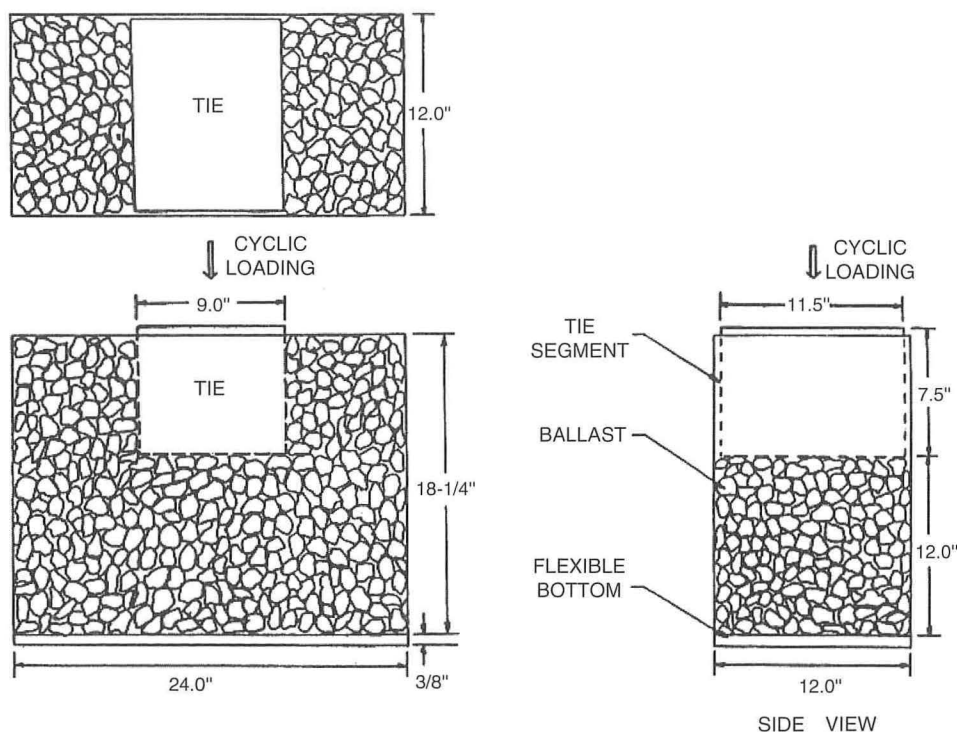


FIGURE 2 Schematic of box test apparatus taken from Chiang (11), as adapted from Roner and Selig (1984).

## DEVELOPMENT OF FINITE ELEMENT MODEL

### Wood Box Model

The box test apparatus confines the ballast at either end of the tie section, and the tie is very stiff in comparison to the ballast. A two-dimensional plane strain model, therefore, is a good approximation of the ballast behavior under the testing conditions and was chosen for the FE model. The sides of the box were modeled as rigid in the horizontal direction and free to move in the vertical direction. The base of the box was covered with soft foam rubber overlaid by wood boards and was instrumented with both stress- and strain-measuring instruments. To simulate the stress-strain conditions measured at the flexible base and estimate the elastic modulus, an FE model was constructed in which a distributed load equal to the stress measured during testing was placed directly on the base. Poisson's ratio was assumed to be zero because the foam rubber was installed in pieces with spaces between, which would allow for free horizontal expansion. The elastic modulus for the base layer was adjusted in the FE model until the strain matched that observed in the box test, resulting in an elastic modulus of 220 psi (1.52 MPa).

A Mohr-Coulomb elastic-plastic soil model, in which linear elastic and perfectly plastic behavior is assumed, was chosen for the material model because it is well suited for granular materials such as gravel. The shear strength is taken as

$$\tau = c' + \sigma_v' \tan \phi' \quad (1)$$

where

$\tau$  = shear strength,

$c'$  = cohesion in units of stress,

$\sigma'_v$  = vertical effective stress, and  
 $\phi'$  = angle of internal friction.

Values of  $\phi'$  ( $45^\circ$ ) and  $c'$  (0.1 psi [0.69 kPa]) were chosen as typical values for ballast material. Clean ballast does not have any cohesive strength; however, the PLAXIS FE code does not work well with zero cohesive strength (13), so a minimal nonzero value was assumed. The elastic modulus was then calibrated from the box test deflection results to be 26,800 psi (184.8 MPa), which is near typical values for the material under the loading conditions of 27,900 psi (192.4 MPa) (5).

An ultimate capacity analysis of the loaded railroad tie was performed to evaluate the behavior of the system at the upper limit of its capacity. This analysis is performed by incrementally reducing  $\phi'$  and  $c'$  simultaneously until the ballast body collapses. The box test was modeled with and without the flexible base. Figure 3 presents the results of both capacity analyses, and Figure 4 shows the mechanisms of plastic failure. The presence of the cushion appears to cause the failure mode to closely resemble punching shear, while the absence of the cushion causes rotational instability. The safety factor, defined as the ratio of the initial shear strength to the reduced shear strength at failure, was found to be 1.24 for the stiff base and 1.15 for the flexible base. The flexible base reduces the ballast bearing capacity, as may be expected. Furthermore, the analysis shows that the tests were performed at states near the ultimate capacity of the ballast body. Figure 5 shows the distribution of plastic points in the FE model just before failure for the flexible base condition.

Figure 6 shows a comparison of the horizontal stresses at the boundaries of the FE model with those measured at the sides and ends of the box. The end stresses compare well except for the one measured at the box just below the tie, which drops off inexplicably; the FE stress on the end below the tie is consistent with expectations in that the stress just below the tie is nearly equal to that applied to it. The horizontal stresses at the sides of each model appear to mirror each other. The FE distribution is again consistent with expectations because the stress should spread away from the tie with depth. The opposite trend observed with the box test model is difficult to explain in the context of continuum mechanics. It may be attributable to the reaction force at the bottom of the box being concentrated beneath the tie, that is, there may not be enough depth below the tie to allow

for even stress distribution. The punching shear failure mechanism exhibited in the FE model supports this hypothesis.

### Steel Box Model

Another FE model was developed for the steel box test. The steel box had a firm rubber lining on both the bottom and sides covered with a thin sheet of stainless steel, but it was not instrumented for stress or strain measurements. A typical value of an elastic modulus for firm rubber was chosen to be 7,500 psi (51.7 MPa). The addition of the rubber to the sides of the FE model produced rotational instability in the model (Figure 7).

The deflections obtained in the steel box test were not reproducible by the FE model (using reasonable soil property values for the ballast) without plastic failure and collapse of the soil body. The box test deflections may, therefore, have had a large plastic deformation component caused by the low friction angle (approximately  $24^\circ$ ) of the smooth steel walls in combination with the flexible rubber sides. In order to simulate the steel box test, the FE model was increased in size to 48 in. (142 cm) wide by 36 in. (91 cm) high, and the flexible rubber boundary conditions were removed. The larger simulated box was able to produce deflections equal to the original box test data with reasonable ballast material properties. Although the larger size of the FE model is a departure from the box test model geometry, it is an approximation that is meant to account for the boundary conditions of the box test (which in turn were intended to simulate the actual conditions beneath a railroad tie). Figure 8 shows a comparison of the shear stress distributions between the large box and the small box. The increased size of the large box eliminated the shear stress concentrations evident in the small box, which had caused plastic failure of the soil body.

### Calibration of Model

The steel box FE model was calibrated with respect to percentage fouling of clay at its liquid limit (85% water content) with the tie deflection results of the box tests. Panther Creek clay was used as the fouling material, with a liquid limit of 85% and a plastic limit of 46%. The list below gives the properties of the clean ballast material, Dell

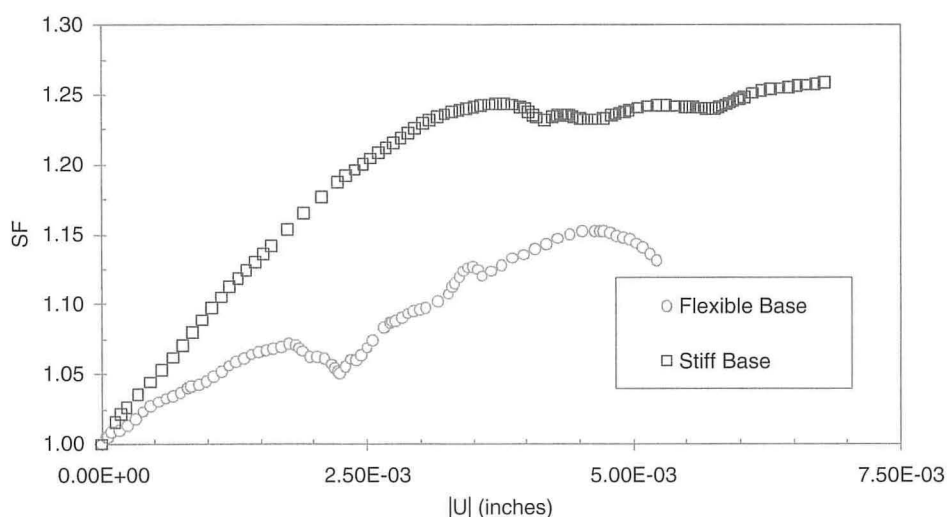


FIGURE 3 Ultimate results ( $|U|$ ) of capacity analysis (SF = safety factor).

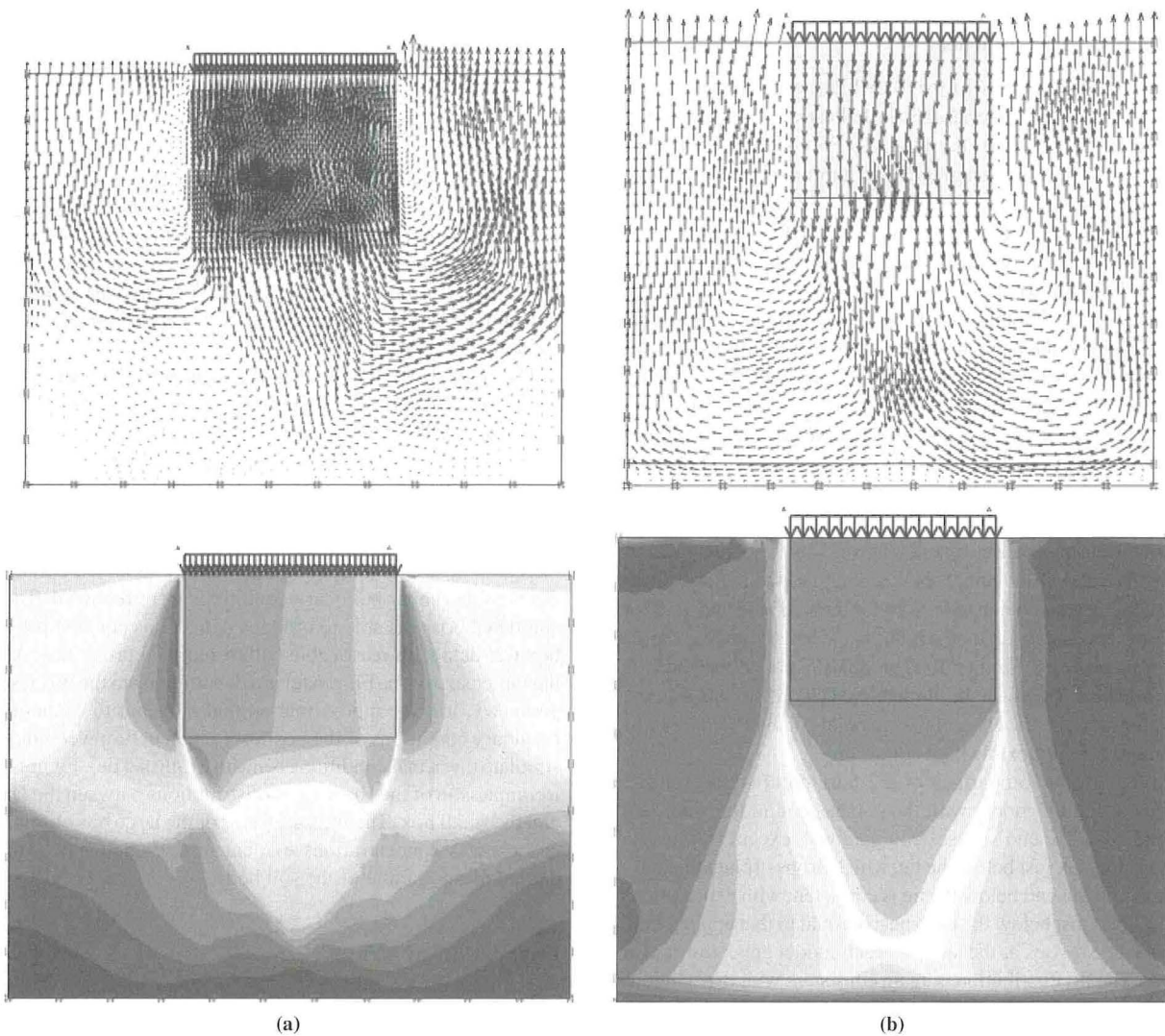


FIGURE 4 Comparison of failure modes (a) with and (b) without flexible base showing paths of (top) plastic flow and (bottom) deformation gradients.

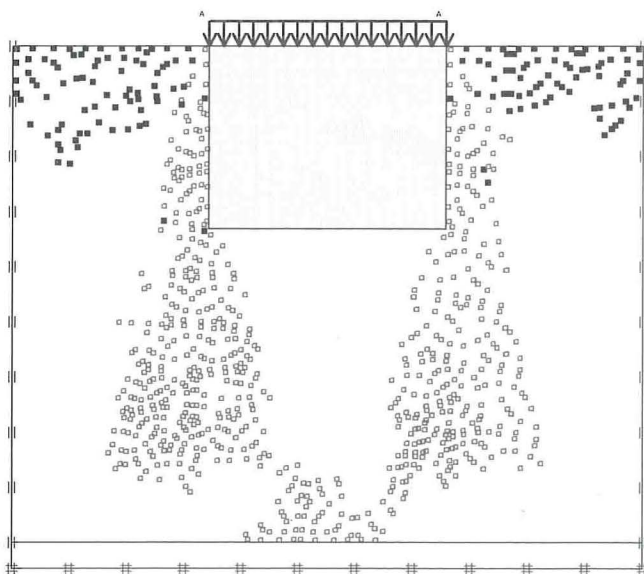


FIGURE 5 Plastic points at failure (light gray) in flexible base capacity analysis.

Rapids quartzite, which was described as “fairly rounded” after being used for approximately 3 years in track (8):

- Specific gravity: 2.66;
- Porosity,  $n$ : 0.37;
- Dry unit weight,  $\gamma_d$  (kN/m<sup>3</sup>): 16.4;
- Particle diameter that 60% of ballast by mass is finer than 19.5 mm;
- Particle diameter that 10% of ballast by mass is finer than 37.5 mm;
- Coefficient of uniformity: 1.92; and
- Coefficient of curvature = 1.00.

Typical values of  $\phi'$  and  $c'$  for gravel–clay mixtures are difficult to define because of the variability involved, so judgment had to be used in evaluating available data and to increment changes in these strength parameters with respect to percentage fouling. Figure 9 presents a comparison of the friction angles chosen for this study with friction angles measured in laboratory direct shear tests on ballast fouled with clay (14). The effective cohesion ( $c'$ ) was nominally increased with percentage fouling.

The elastic modulus ( $E$ ) as a function of percentage fouling was established for the model and is presented in Figure 10. The result-

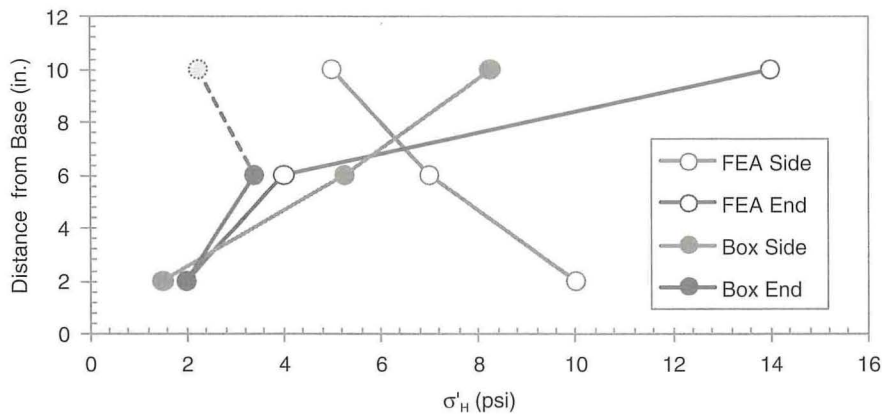


FIGURE 6 Comparison of horizontal stresses ( $\sigma'_H$ ) in box test and FE models.

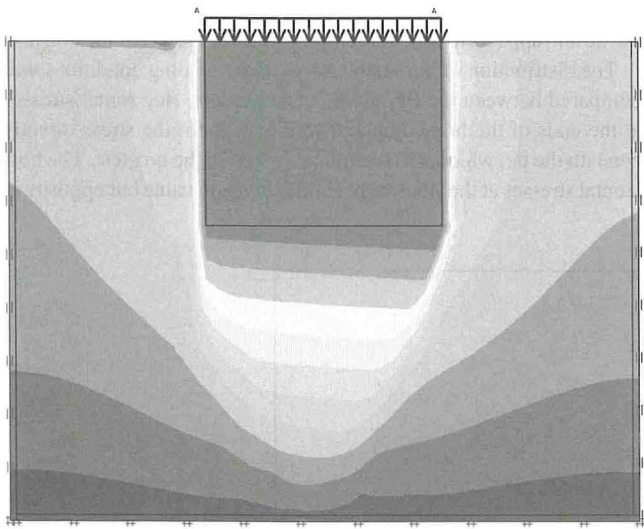


FIGURE 7 Deformation gradient showing rotational instability caused by rubber sides in steel box.

ing elastic modulus trend decreases linearly until about 20% fouling and then flattens out, even increasing slightly with increased fouling after this point. The early decreasing trend can be explained as the result of decreased friction caused by lubrication of the ballast by the clay, being actually the manifestation of increased plastic deformation. The stability and even slight increase in elastic modulus beyond 20% fouling could mean that the composite strength parameters of the soil body are no longer decreasing, but further laboratory investigation is required to examine this behavior.

Figure 11 presents the simulated deflection curves for various degrees of fouling, and Table 1 summarizes the results of the fouled ballast FE model calibration. The ultimate deflections were taken from the box test results (8) as the elastic portion of the measured deflection from a single load after the first 1,000 cycles of a loading test. The loads after the first 1,000 cycles were chosen because the rate of plastic settlement was observed to generally reach a minimum and become constant at that point (8). The FE model simulation of the plastic component of the deformation can be seen as curvature in Figure 11; however, these characteristics are model simulations and do not necessarily capture the true behavior of the box test because the actual strength parameters of the fouled ballast were not measured

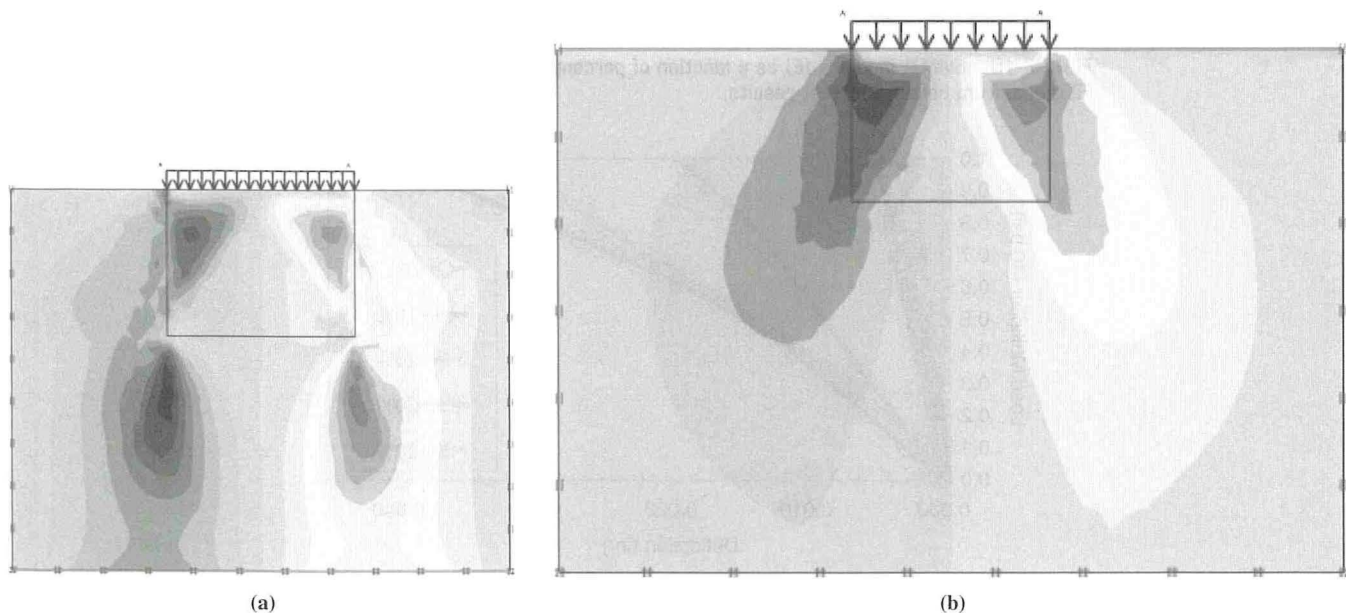


FIGURE 8 Comparison showing shear concentration in (a) small box absent in (b) large box.

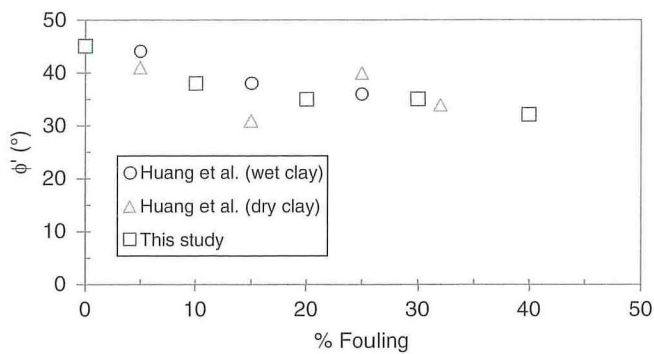


FIGURE 9 Comparison of friction angles ( $\phi'$ ) of ballast chosen for FE model with data from Huang et al. (14).

at the time of the box tests. Further box tests on fouled ballast made in conjunction with triaxial tests would allow for greater confidence in characterizing the plastic behavior. The box test measurements for the 40% fouling test were extrapolated beyond 100 cycles and exhibited the highest degree of plastic deformation (8). The high degree of plastic deformation resulted in a lower elastic deformation component interpreted from the box test results, causing the 40% fouling test to be an exception to the observed trend of increasing elastic

deformation with increased fouling. Further testing is required to confirm the extrapolated test results for 40% fouling.

## CONCLUSION

The FE model is able to simulate box test results and provides insight into the mechanisms of ballast–apparatus interaction. The conditions at the boundaries of the box (i.e., stiffness, smoothness, and size) appear to play a large role in the stress distribution and plastic behavior of the ballast. The addition of flexible and smooth material at the sides has been shown to significantly reduce the strength capacity of the ballast, increase plastic deformation, and introduce rotational instability of the tie. A flexible boundary located only at the base does not introduce rotational instability; however, it does cause the failure mode to resemble punching shear and reduces the overall capacity of the ballast. An increase in box dimensions of approximately 50% was sufficient to avoid shear stress concentrations below the tie and appears to account for the boundary effects of the box test.

The distribution of horizontal stress under loading conditions was compared between the FE and box test models. Horizontal stresses at the ends of the box compared well except for the stress directly beneath the tie, which was inexplicably low in the box test. The horizontal stresses at the sides were similar in mean value but opposite in

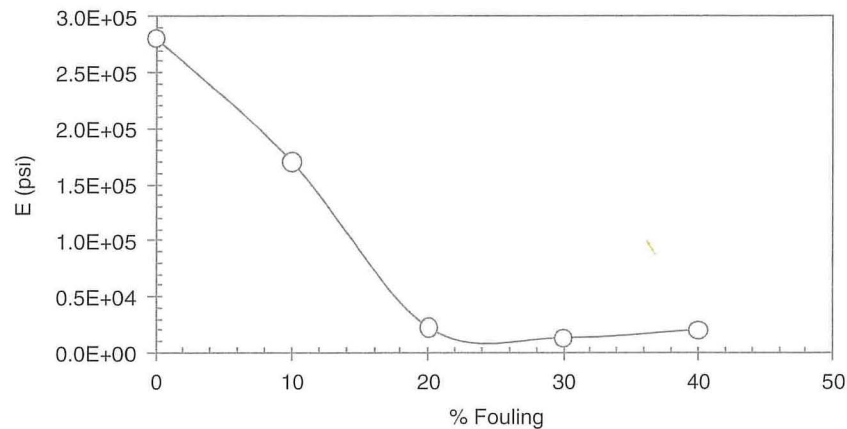


FIGURE 10 Elastic modulus ( $E$ ) as a function of percentage fouling as calibrated by FE model from ballast box test results.

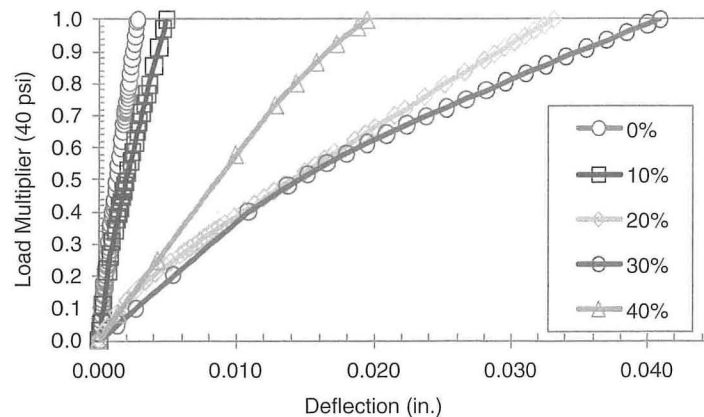


FIGURE 11 FE model tie deflection curves for fouled ballast with increment of load on y-axis.

TABLE 1 Fouled Ballast Parameters of FE Model

Fouling (%)	$\delta$ (in.)	$c'$ (psi)	$\phi'$ (°)	$E$ (psi)
0	0.029	0.05	45	2.80 E+05
10	0.031	0.10	38	1.70 E+05
20	0.060	0.50	35	2.20 E+04
30	0.070	2.00	35	1.28 E+04
40	0.046	4.00	32	2.00 E+04

NOTE:  $\delta$  = vertical deflection.

trend, which may point to inaccuracies in the box test model resulting from insufficient depth of ballast below the tie.

The developed FE model has been calibrated for modeling the behavior of ballast at various degrees of fouling with clay. The relationship of the elastic modulus to percentage fouling has been established and can be used for future comparisons and models. Fouled ballast parameters vary considerably and were estimated for the model.

This study has suggested two areas of future work that may improve the accuracy of both the box test and FE model: (a) triaxial testing of fouled ballast in conjunction with box tests and (b) construction of a larger box (or multiple larger boxes) for comparison testing. Triaxial test data would provide strength parameters for the FE model, improving its accuracy and enabling better examination of the plastic deformation behavior of the ballast. Boxes of various sizes would allow for a study of scale effects and may provide information useful to optimization of the box test design.

## ACKNOWLEDGMENT

This material is based on work supported by the National Science Foundation.

## REFERENCES

1. Ho, C. L., J. P. Hyslip, and D. D. Davis. Ground Acceleration-Based Characterization of High-Speed Rail Track Bed. *Proc., ASCE Geo-Congress 2008*, American Society of Civil Engineers, New Orleans, La., 2008.
2. Lambert, R. S. *An Inquiry into the Degradation and Fouling of Railway Ballast*. BS honors report. University of Massachusetts, Amherst, 2009.
3. Lambert, R. S., C. L. Ho, and J. P. Hyslip. Time and Spatial Dependence of the Fouling of Railroad Track Ballast. *GeoFrontiers 2011*, Geo-Institute, American Society of Civil Engineers, Dallas, Tex., 2011.
4. Zarembski, A. M., and J. F. Cikota, Jr. Estimating Maintenance Costs for Mixed High-Speed Passenger and Freight Rail Corridors: A New Tool for Rail Planners. *TR News*, No. 255, March–April 2008, pp. 29–31.
5. Selig, E. T., and J. M. Waters. *Track Geotechnology and Substructure Management*. Thomas Telford Services Ltd., London, 1994.
6. HAL Revenue Service Testing Update: Eastern Mega Site. *Railway, Track and Structures*, July 2006, pp. 15–17.
7. Norman, G. M. *Ballast Box Experiments for Evaluating Ballast Field Performance*. Concrete tie correlation study Project Report FRA82-291P. University of Massachusetts, Amherst, 1982.
8. Han, X., and E. T. Selig. *Investigation of the Effects of Fouling Material and Degree of Fouling on the Settlement of Ballast Bed by Ballast Box*. Project Report No. AAR95-426R for Association of American Railroads. University of Massachusetts, Amherst, 1996.
9. McDowell, G. R., W. L. Lim, A. C. Collop, R. Armitage, and N. H. Thom. Comparison of Ballast Index Tests for Railway Trackbeds. *Proceedings of the Institution of Civil Engineers: Geotechnical Engineering*, Vol. 157, No. 3, 2004, pp. 151–161.
10. McDowell, G. R., W. L. Lim, A. C. Collop, R. Armitage, and N. H. Thom. Laboratory Simulation of Train Loading and Tamping on Ballast. *Proceedings of the Institution of Civil Engineers: Transport*, Vol. 158, No. 2, 2005, pp. 89–95.
11. Chiang, C. C. *Effects of Water and Fines on Ballast Performance in Box Tests*. Project Report No. AAR89-366P for Association of American Railroads. University of Massachusetts, Amherst, 1989.
12. Lu, M., and G. R. McDowell. The Importance of Modeling Ballast Particle Shape in the Discrete Element Method. *Granular Matter*, Vol. 9, No. 1–2, 2007, pp. 69–80.
13. R. B. J. Brinkgreve (ed.). *PLAXIS: Version 8, Material Models Manual*. PLAXIS, Delft, Netherlands, 2002.
14. Huang, H., E. Tutumluer, and W. Dombrow. Laboratory Characterization of Fouled Railroad Ballast Behavior. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2117, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 93–101.

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The Railway Maintenance Committee peer-reviewed this paper.

# Analysis of Derailments by Accident Cause

## Evaluating Railroad Track Upgrades to Reduce Transportation Risk

Xiang Liu, Christopher P. L. Barkan, and M. Rapik Saat

The risk of train derailment associated with rail transportation is an ongoing concern for the rail industry, government, and the public. Various approaches have been considered or adopted to analyze, manage, and reduce risk. Upgrading track quality has been identified as one possible strategy for preventing derailment. The quality of freight railroad track is commonly divided into five principal classes by FRA on the basis of track structure, track geometry, and inspection frequency and method. The higher the track class, the more stringent are the track safety standards and thus a higher maximum train speed is allowed. Upgrading track class is likely to prevent certain track-related derailments; however, this upgrade may also increase the risk of certain types of equipment failure that are more likely to occur at higher speeds. Consequently, more sophisticated approaches need to be developed to examine the interactions among accident causes that may be differently affected by upgrades to track infrastructure. This paper analyzes several critical parameters for predicting train derailment risk by using derailment statistics from the FRA accident database and related literature. A general method was developed to assess derailment risk by accident cause and FRA track class. The safety benefits of track class upgrade in reducing the risks from certain accident causes were quantitatively evaluated. The model can be extended by incorporating additional risk factors to more accurately assess the effectiveness of various derailment prevention efforts for reducing transportation risk.

Train derailments are the most common type of main-line railroad accident in the United States, causing property damage and service disruptions. Reducing derailment risk is an ongoing objective of the rail industry and government. Derailment risk analysis provides a scientific basis for evaluating various risk reduction options. It is concerned with derailment rate, which reflects the likelihood that a train is involved in a derailment, and the consequences of the derailment. FRA requires reporting and identification of specific accident cause(s) for all derailments that exceed a specified monetary damage threshold. The monetary threshold is periodically adjusted for inflation; in 2010, the reporting threshold was \$9,200. Different accident causes may have different effects on derailment rate, derailment severity, and the corresponding risk for any given combination of infra-

structure and operational conditions. Track-related and equipment-related accident causes collectively result in the majority of train derailments in the United States. Understanding how they may be affected by approaches to derailment prevention is useful for developing and evaluating cost-effective strategies to reduce railroad transportation risk. Upgrading track quality is one possible derailment prevention strategy. FRA divides track quality into five principal classes commonly used by freight railroads in accordance with FRA track safety standards. Higher track classes have correspondingly higher maximum train speeds and more stringent track safety standards (1).

Research has analyzed various safety and economic impacts of track class upgrade. Higher track classes are statistically correlated with lower derailment rates (2–4). Saat and Barkan developed an analytical model to compare the safety benefits of enhanced tank car safety design versus infrastructure improvement (5). Lai et al. developed an optimization framework to determine optimal track class assignment based on the minimization of track maintenance and transportation costs (6). Liu et al. presented a benefit–cost analysis framework to consider the trade-off between reduced accident rates and increased track maintenance costs in evaluating track class upgrade as a risk reduction strategy (7). Kawprasert proposed a biobjective model that simultaneously considers risk and investment costs in determining the best track infrastructure upgrade strategy (8). However, none of these investigations addressed how track class upgrade affects the risk pertaining to certain accident causes. Although upgrading track class is expected to prevent certain track-related derailments, it may also increase the risks from certain types of equipment failure that are more likely to occur at higher speeds.

This study developed an accident cause–specific derailment risk model that simultaneously accounts for the interactions among different accident causes that may be differently affected by track class upgrade. The paper is structured as follows: a general framework for derailment risk analysis is introduced, followed by analyses and modeling of derailment rate, severity, and the corresponding risks. Finally, accident cause–specific derailment risk by FRA track class is estimated using derailment statistics from the FRA Accident/Incident Reporting System database and recent literature.

### FRAMEWORK FOR DERAILMENT RISK ANALYSIS

Risk has been defined as the function of system failure and the severity of losses or damages from the system failure (9). In the context of railroad transportation, train derailment risk is defined as a product of derailment frequency and the average consequences of the derail-

Rail Transportation and Engineering Center, Department of Civil and Environmental Engineering, University of Illinois at Urbana–Champaign, 205 North Mathews Avenue, Urbana, IL 61801. Corresponding author: X. Liu, liu94@illinois.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 178–185.  
DOI: 10.3141/2261-21

ment. Derailment frequency is a product of derailment rate and traffic exposure (2, 10–13). Equation 1 is a general framework for derailment risk analysis:

$$R = Z \times M \times D \quad (1)$$

where

- $R$  = derailment risk,
- $Z$  = derailment rate,
- $M$  = traffic exposure, and
- $D$  = average consequences of a derailment.

Other than traffic volume, the two probability terms of most importance are derailment rate and severity. Derailment rate is a critical metric to measure railroad transportation safety performance. It is defined as number of derailments normalized by some measure of traffic exposure, that is, gross ton-miles, car miles, or train miles. Derailment occurrence can be modeled as a Poisson process in which the Poisson parameter is numerically equal to derailment rate (8, 14–16). The probability of  $k$  derailments within a traffic interval  $M$  can be expressed as follows:

$$P\{N(M) = k\} = \frac{\exp(-ZM)(ZM)^k}{k!} \quad (2)$$

where

- $P\{N(M) = k\}$  = probability of  $k$  derailments for traffic exposure  $M$ ;
- $N(M)$  = number of derailments for traffic exposure  $M$ ;
- $k = 0, 1, 2, \dots$ ; and
- $Z$  = derailment rate (train derailments per billion car miles).

Consequently, the expected number of derailments within traffic level  $M$  can be estimated using Equation 3, which shows that predicted derailment frequency is a product of derailment rate and traffic exposure:

$$\begin{aligned} E[N(M)] &= \sum_k P[N(M) = k] \times k \\ &= \sum_k \left\{ \frac{\exp(-ZM)(ZM)^k}{k!} \right\} \times k \\ &= ZM \end{aligned} \quad (3)$$

Many factors demonstrate a correlation with derailment rates, including FRA track class (2–4, 10), type of track and railroad (17), train length (18, 19), track geometry (20, 21), and train control system (22, 23). Train accident rates in the United States have declined substantially since the early 1980s as a result of major capital investments in infrastructure and equipment, employee training, and implementation of new technologies (17, 24). Derailment severity also varies widely, depending on train speed, infrastructure conditions, and derailment cause. When risk analysis is concerned with the release of hazardous materials, the general risk model can be extended by adding a series of possible release consequences and associated probabilities.

The objective of this research was to analyze several critical factors that affect derailment risk analysis. The study focused on main-line derailments on Class I railroads (gross annual revenue in 2006  $\geq$  \$346.8 million) using data from the FRA accident database from 1999 to 2008.

## CAUSE-SPECIFIC DERAILMENT RATE

Higher track classes are statistically correlated with lower derailment rates (2–4). Anderson and Barkan used FRA safety statistics to develop the most recent published estimates of main-line freight train derailment rates for Class I railroads (Table 1) (4). Their estimates are used in this paper to account for the rates of all main-line derailments on Class I railroads. It was assumed for this study that the overall Class I main-line derailment rates did not change significantly during the study period. However, an approach that allows for varying future derailment rates could be developed for a longer-term analysis.

FRA requires the identification of a primary accident cause (and other contributing causes if applicable). These causes are categorized into five major groups in the FRA accident database: track, equipment, human factors, signals, and miscellaneous. Track-related causes and equipment-related causes collectively account for the majority of derailments and are the focus of this research. A three-digit numeric code was assigned to an individual cause that specifies the exact cause(s) of an accident. Different accident causes may be associated with different track class-specific derailment rates. In addition, some causes may have no relationship or only an indirect relationship with track class. It is of interest to analyze accident cause-specific derailment rates by FRA track class. The FRA accident database contains 389 unique accident causes (25, 26). A study by Arthur D. Little Inc. combined similar FRA causes into the same cause group on the basis of expert opinion (27). Accident cause-specific derailment rates by FRA track class were estimated using the overall derailment rates for all causes developed by Anderson and Barkan (4), multiplied by the conditional probability that a derailment is the result of a certain accident cause (Equations 4 and 5). This approach assumed that the distribution of accident cause on each track class remained constant in the study period. In other words, the conditional probability that a derailment is the result of a certain cause was held constant. An alternative method is to divide the number of derailments that resulted from a certain accident cause by the corresponding track class-specific traffic exposure. This method requires information regarding freight rail traffic distribution by track class from 1999 to 2008, which is not publicly available.

$$Z_{ck} = Z_{ak} \times P_{ck|ak} \quad (4)$$

$$P_{ck|ak} = \frac{F_{ck}}{F_{ak}} \quad (5)$$

TABLE 1 Derailment Rates by FRA Track Class for Class I Main-Line Freight Trains (4)

Class	Train Derailments per Million Freight Train Miles	Train Derailments per Billion Freight Car Miles
X and 1	48.54	720.1
2	6.06	92.7
3	2.04	31.5
4	0.53	7.8
5	0.32	4.9
All classes	1.00	14.8

where

$Z_{ck}$  = derailment rate from cause  $c$  on track class  $k$  (per billion car miles),

$Z_{ak}$  = derailment rate from all causes on track class  $k$  (per billion car miles),

$P_{ck|ak}$  = conditional probability that a derailment on track class  $k$  is from cause  $c$ ,

$F_{ck}$  = number of derailments from cause  $c$  on track class  $k$ ,

$F_{ak}$  = number of derailments from all causes on track class  $k$ ,

$c$  = accident cause, and

$k$  = track Class 1, 2, 3, 4, or 5.

In general, track-related derailments are more frequent than equipment-related derailments on lower track classes, while equipment-related derailments are more frequent than track-related derailments on higher track classes (Figure 1).

Table 2 presents Class I main-line train derailment rates by track class and accident cause. Several observations can be made:

- Broken rails or welds have the highest derailment rate on each track class among all track-related ( $T$ ) and equipment-related ( $E$ ) causes;
- Bearing failures and broken wheels have higher derailment rates than the other equipment-related causes on track Classes 3, 4, and 5; side bearing and suspension defects have higher derailment rates than the other equipment-related causes on track Classes 1 and 2; and
- Broken rails or welds (08T), bearing failures (10E), track geometry defects (04T), and broken wheels (12E) are the principal track-related and equipment-related causes.

An analysis of variance (ANOVA) was conducted to examine the effects of FRA track class and accident cause on derailment rate. The  $p$ -value for FRA track class was less than .05 for both track cause group and equipment cause group, indicating that FRA track class is significantly related to derailment rate. The different track-related causes did not have a significantly different relationship in terms of track class ( $p = .202$ ). However, the relationship for equipment-related causes did differ significantly with track class ( $p < .05$ ). The results of the ANOVA show that FRA track class is a significant fac-

tor affecting derailment rates and that the relationship with accident causes varies depending on major cause group. To better understand the relationship between derailment rate and FRA track class for specific accident causes, a Pearson product-moment correlation coefficient was calculated. It was assumed that there is no significant linear relationship between derailment rate and track class if  $|r| < .5$ ; otherwise, a linear correlation is expected (28). A positive  $r$  value indicates that derailment rate increases with a higher track class; a negative  $r$  value means that a higher track class is correlated with a lower derailment rate. In general, higher track classes were correlated with lower derailment rates for all track-related causes and the majority of equipment-related causes. However, two equipment-related causes, truck structure defects (08E) and especially hunting (20E), showed a positive relationship between higher track class and derailment rate. Some equipment-related causes, such as bearing failure (10E) and air hose defect (01E), had no relationship with track class.

It is interesting to observe that some equipment-related accident causes, such as broken wheels (12E), also have lower derailment rates on higher track classes. This might be attributable to the reduced dynamic forces on the vehicle components resulting from the higher track quality. In addition, installation of advanced wayside detectors (29–35), such as wheel impact load detectors, may result in certain equipment defects being more likely to be identified and corrected before causing a derailment. Sampling error may also be a factor when there are few derailment records for a particular cause group.

The differences in derailment rates by track class are not solely explained by track quality variations in track classes. Some derailments may occur on curves with lower allowable speeds or on sections with temporary slow orders. These segments are likely to be classified as lower FRA track classes; however, when such circumstances occur on high-density main-line routes that are otherwise being maintained for higher track class standards, most of the same maintenance standards will prevail. The effect will be to reduce the estimated derailment rates on these lower track class segments. Additionally, railroads may maintain their infrastructure to a higher standard than the minimum required by FRA, thereby causing possible variations in track quality within the same track class. FRA has developed a set of objective track quality indices from measured track

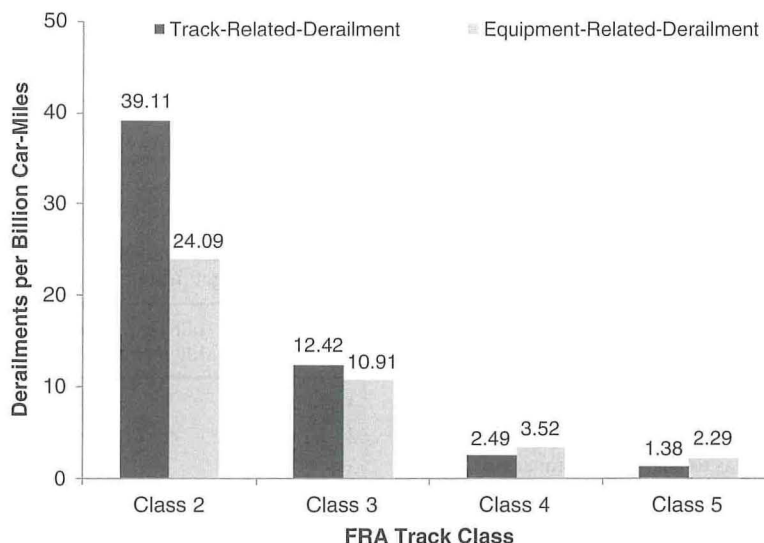


FIGURE 1 Track-related and equipment-related derailment rates.

TABLE 2 Derailment Rates by FRA Track Class, Class I Main-Line Freight Train, 1999 to 2008 (Derailments per Billion Car Miles)

Cause Group	Description	Class 1	Class 2	Class 3	Class 4	Class 5	All Classes	Correlation Coefficient with Track Class, <i>r</i>
Track Related								
08T	Broken rails or welds	129.07	16.78	5.01	0.94	0.59	2.17	-.78
04T	Track geometry (excl. wide gauge)	61.14	10.01	2.43	0.40	0.07	0.99	-.79
03T	Wide gage	81.52	4.06	0.66	0.10	0.02	0.48	-.73
05T	Buckled track	6.79	1.89	1.14	0.33	0.09	0.45	-.86
10T	Turnout defects: switches	44.84	1.49	0.59	0.12	0.08	0.33	-.73
07T	Joint bar defects	2.72	0.41	0.48	0.18	0.15	0.24	-.78
12T	Miscellaneous track and structure defects	31.25	1.62	0.26	0.07	0.06	0.22	-.73
01T	Roadbed defects	17.66	0.81	0.63	0.10	0.02	0.21	-.74
09T	Other rail and joint defects	4.08	0.95	0.41	0.10	0.05	0.16	-.83
06T	Rail defects at bolted joint	4.08	0.27	0.22	0.09	0.18	0.16	-.72
02T	Nontraffic, weather causes	5.43	0.68	0.48	0.05	0.02	0.12	-.78
11T	Turnout defects: frogs	1.36	0.14	0.11	0.03	0.04	0.05	-.76
Total		389.94	39.11	12.42	2.49	1.38	5.59	-.76
Equipment Related								
10E	Bearing failure (car)	0.00	1.89	1.99	0.85	0.46	1.01	-.02
12E	Broken wheels (car)	5.43	2.03	2.03	0.51	0.46	0.78	-.90
13E	Other wheel defects (car)	10.87	3.52	0.99	0.31	0.11	0.48	-.87
11E	Other axle, journal defects (car)	4.08	1.62	0.85	0.31	0.24	0.44	-.90
09E	Sidebearing, suspension defects	13.59	4.06	1.03	0.21	0.08	0.43	-.86
07E	Coupler defects (car)	2.72	1.22	0.77	0.37	0.08	0.41	-.94
06E	Centerplate, car body defects (car)	5.43	1.89	0.88	0.18	0.14	0.33	-.88
Loco	Locomotive defects	8.15	1.62	0.44	0.16	0.19	0.28	-.80
19E	Stiff truck (car)	5.43	3.25	0.41	0.08	0.04	0.21	-.91
20E	Track-train interaction (hunting)	0.00	0.00	0.07	0.15	0.18	0.17	.97
18E	All other car defects	6.79	0.95	0.37	0.07	0.08	0.15	-.78
05E	Other brake defects (car)	0.00	0.81	0.44	0.08	0.04	0.14	-.30
08E	Truck structure defects (car)	0.00	0.00	0.22	0.09	0.12	0.12	.57
02E	Brake rigging defects (car)	2.72	0.54	0.22	0.07	0.06	0.11	-.81
01E	Air hose defects (car)	0.00	0.14	0.15	0.05	0.01	0.05	-.16
Unclassed	FRA causes not classified by ADL	2.72	0.41	0.04	0.02	0.01	0.04	-.78
04E	UDE (car or locomotive)	0.00	0.00	0.00	0.02	0.00	0.01	.35
14E	TOFC, COFC defects	0.00	0.14	0.00	0.01	0.00	0.01	-.33
03E	Hand brake defects (car)	1.36	0.00	0.00	0.00	0.00	0.00	-.71
Total		69.29	24.09	10.91	3.52	2.29	5.20	-.88

NOTE: Excl. = excluding; ADL = Arthur D. Little Inc.; UDE = undesired emergency (brake application); TOFC = trailer on flat car; COFC = container on flat car. For track-related cause group,  $p = .202$  ( $>.05$ ); for equipment-related caused group,  $p = .020$  ( $<.05$ ); for track class (both groups),  $p = .000$  ( $<.05$ ).

geometry data that can quantitatively describe the relative condition of quality within each track class (36). Additional factors related to track quality, track geometry, and environmental conditions should also be considered for a more accurate estimation of derailment rate.

## DERAILMENT SEVERITY

Monetary damage is often used to assess the severity of railroad accidents. However, this is subject to additional sources of variance such as the cost to repair regular track versus special track and the difference in cost between damaged locomotives and freight cars (17). In

this paper, number of cars derailed is used as a proxy for derailment severity to reflect the impact forces resulting from the kinetic energy of the derailed train. A positive correlation between number of cars derailed and speed is expected (2, 3, 11, 17, 37). The number of cars derailed may also vary among different accident causes. For instance, derailments caused by broken rails or welds tend to derail more cars than those caused by bearing failure (11, 17). Average number of cars derailed can be modeled as a function of speed and accident cause by using a power regression model (2, 11, 37) (Equation 6). The estimates for parameters  $A_c$  and  $B_c$  are presented in Table 3.

$$N_c = A_c \times S^{B_c} \quad (6)$$

TABLE 3 Regression Results for Estimating Average Number of Cars Derailed

Cause Group	Description	$A_c$	$B_c$	$R^2$	$p$
08T	Broken rails or welds	1.830	0.622	.986	<.01
04T	Track geometry (excl. wide gage)	2.952	0.257	.911	<.01
All T	All track causes combined	1.528	0.636	.982	<.01
10E	Bearing failure (car)	3.012	0.196	.497	<.01
12E	Broken wheels (car)	1.265	0.506	.917	<.01
All E	All equipment causes combined	1.665	0.397	.985	<.01
All	All accident causes	1.852	0.486	.981	<.01

where

$N_c$  = average number of cars derailed in a train derailment for accident cause  $c$ ,

$A_c$ ,  $B_c$  = model coefficients for accident cause  $c$ , and

$S$  = train derailment speed (mph).

The regression relationship between average number of cars derailed and derailment speed by accident cause is plotted in Figure 2, which shows that

- Derailments with higher train speeds tend to have more cars derailed within the same cause group.
- The average number of cars derailed varies by accident causes even at the same derailment speed:
  - Broken rails or welds are likely to derail more cars than other causes and
  - Bearing failures result in relatively less severe derailments.

It appears that track-related accident causes result in a larger number of cars derailed than equipment-related causes. This may be because different accident causes vary in the typical point of derailment (POD) (the position of the first car that is derailed). Track-related accident causes, such as broken rails, tend to initiate derailments near the front of the train; equipment failures appear to have the POD more uniformly distributed throughout the train (11). When the POD is located near the front, the larger residual train length (number of

cars behind the POD), would result in more cars derailed. Understanding that different accident causes may result in different numbers of cars derailed is useful for analyzing the risk pertaining to a specific accident cause.

Train length and car loading were not considered in the risk model since this study analyzed accident cause-specific derailment risk by FRA track class. No significant differences have been found in average train length and average car weight between track classes on the national freight rail network (11). But train length and train weight are likely to affect derailment rate, derailment severity, and the corresponding risk in a route-specific risk analysis (11, 19). In that case, more operational and infrastructure information should be incorporated into the model.

## DERAILMENT SPEED

Higher track classes allow higher maximum train-operating speeds, which leads to correspondingly higher average derailment speeds. Nevertheless, these speeds were lower than the maximum allowable speeds, especially on higher track classes (Table 4). Higher track classes were also shown to have a larger standard deviation for derailment speed, probably because higher track class derailments occur at a wider range of speeds. This larger standard deviation implies a greater range of derailment severity on higher track classes, making the prediction of number of cars derailed subject to greater uncertainty.

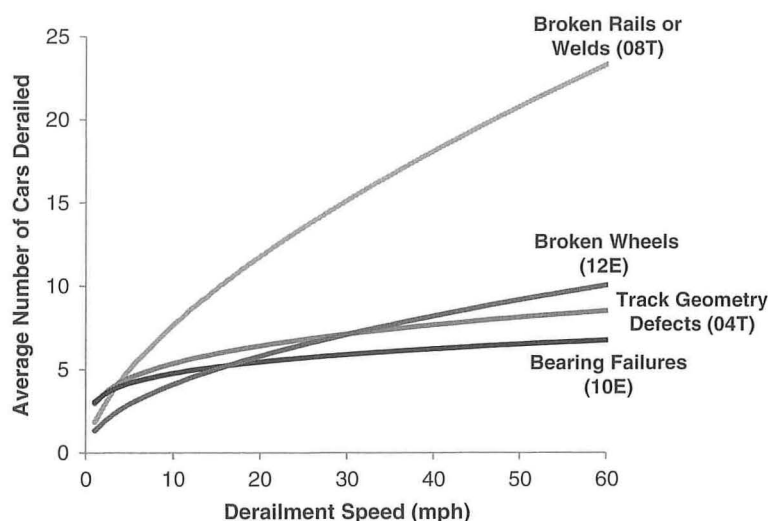


FIGURE 2 Regression relationship between number of cars derailed and speed by accident cause.

**TABLE 4 Average Derailment Speed by Track Class: All Accident Causes**

Class	Number of Derailments	Average Speed (mph)	Standard Deviation (mph)	Freight Train Maximum Speed (mph)
1	530	7.7	3.0	10
2	685	16.9	7.1	25
3	855	25.0	11.2	40
4	1,551	33.2	15.5	60
5	416	37.4	18.0	80
All classes	4,037	25.7	15.9	

Table 5 illustrates the average derailment speed by FRA track class and accident cause. ANOVA indicated that different accident causes have a similar statistical relationship between average derailment speed and track class ( $p > .05$ ). Estimates of accident cause-specific derailment severity were based on the regression relationship between number of cars derailed and train speed (Table 3) and the average derailment speed pertaining to a certain cause on each track class (Table 5). Table 6 illustrates the predicted average number of cars derailed by track class for certain accident cause groups. Broken rails or welds (08T) are likely to result in more cars derailed than the other causes on each track class.

### CAUSE-SPECIFIC DERAILMENT RISK

Derailment rates were shown to decline on higher track classes, but derailment severity (average number of cars derailed) may also increase as a result of higher operating speeds. Derailment risk, a

product of derailment rate and severity, is dominated by the former; consequently, higher track classes have lower derailment risk (Figure 3).

Derailment risks from certain accident causes by track class were analyzed (Table 7). Track-related derailment risk declines on higher track classes because the net result of the reduced derailment rate outweighs the increase in derailment severity. Broken rails or welds (08T) pose greater risk than the other causes on each track class, which is why the detection and prevention of broken rails is a high-priority safety activity for U.S. railroads. Although important on low track classes, derailment risk that results from track geometry defects is relatively low on higher track classes, probably because of more stringent track geometry and maintenance standards. The average derailment risk for all track-related causes is approximately two times that of equipment-related causes. Upgrading track Class 3 to Class 4 is correlated with a fivefold reduction in track-related derailment risk. Although the derailment risk caused by bearing failures and broken wheels increases when track Class 2 is upgraded to Class 3, the overall derailment risk for all causes declines on higher track classes.

### UPGRADING TRACK CLASS TO REDUCE RISK

Three track class upgrade strategies were considered: upgrading Class 2 to 3, Class 3 to 4, and Class 4 to 5. It was assumed that maximum train speeds increase in accordance with track class upgrade, indicating that higher track classes have greater average derailment speeds. The statistics from Table 7 were used to analyze the reduction in derailment risk for a certain accident cause group as a result of track class upgrade. Upgrading track Class 2 to Class 3 offers the greatest risk reduction per billion car miles for track-related causes, but this upgrade is also correlated with an increase in risk caused by broken wheels and bearing failures. Upgrading Class 4 to

**TABLE 5 Average Derailment Speed (mph) by Accident Cause and Track Class**

Cause Group	Class 1	Class 2	Class 3	Class 4	Class 5	All Classes
Broken rails or welds (08T)	8.1	18.1	29.1	36.8	36.0	26.5
Track geometry (excl. wide gage) (04T)	8.5	18.2	26.4	32.5	30.5	23.0
All track causes	8.1	17.5	27.3	36.5	39.9	25.8
Bearing failure (10E)	NA	21.6	27.4	36.3	42.1	34.6
Broken wheels (12E)	7.8	17.7	25.2	36.3	38.9	38.9
All equipment causes	7.2	17.8	24.9	34.8	39.5	30.3
All accident causes	7.7	16.9	25.0	33.2	37.4	25.7

NOTE: NA = not applicable. For cause group,  $p = .98 (>.05)$ ; for track class,  $p = .000 (<.05)$ .

**TABLE 6 Estimated Average Number of Cars Derailed by Accident Cause and Track Class**

Cause Group	Class 1	Class 2	Class 3	Class 4	Class 5	All Classes
Broken rails or welds (08T)	6.7	11.1	14.9	17.2	17.0	14.0
Track geometry (excl. wide gage) (04T)	5.1	6.2	6.9	7.2	7.1	6.6
All track causes	5.8	9.4	12.5	15.1	15.9	12.1
Bearing failure (10E)	NA	5.5	5.8	6.1	6.3	6.0
Broken wheels (12E)	3.6	5.4	6.5	7.8	8.1	8.1
All equipment causes	3.6	5.2	6.0	6.8	7.2	6.5
All accident causes	5.0	7.3	8.9	10.2	10.8	9.0

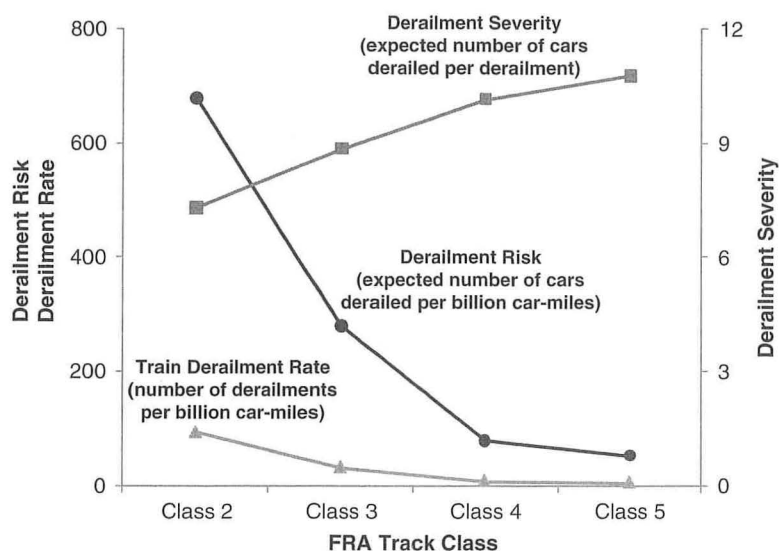


FIGURE 3 Relationship between track class, derailment rate, and derailment severity for all accident causes.

TABLE 7 Derailment Risk by Accident Cause and Track Class, per Billion Car Miles

Cause Group	Class 2	Class 3	Class 4	Class 5	All Classes
Broken rails or welds (08T)	186.3	74.6	16.2	10.0	30.4
Track geometry (excl. wide gage) (04T)	62.1	16.8	2.9	0.5	6.5
All track causes	367.6	155.3	37.6	21.9	67.6
Bearing failure (10E)	10.4	11.5	5.2	2.9	6.1
Broken wheels (12E)	11.0	13.2	4.0	3.7	6.3
All equipment causes	125.3	65.5	23.9	16.5	33.8
All accident causes	676.7	280.4	79.6	52.9	133.2

Class 5 offers less risk reduction than upgrading Class 3 to Class 4 (Table 8).

The derailment risk discussed above is normalized by car miles. Traffic distribution varies among track classes. Track Classes 3, 4, and 5 collectively account for more than 90% of national freight rail traffic, with the majority of traffic on track Class 4 (unpublished data from the University of Illinois). Therefore, track class upgrade between these higher track classes contributes to greater risk reduction. In addition to safety improvement, track class upgrade may also reduce transportation time and enhance rail line capacity. However, upgrading track class also incurs an initial track upgrade cost, and increases operating and ongoing capital costs for track maintenance (7, 38). The trade-offs between the benefits and costs, budget, engineering

constraints, and many other criteria need to be considered when making decisions regarding infrastructure improvement for a specific rail route or network.

## CONCLUSIONS

Train derailment risk analysis relies on the accurate estimation of derailment rate and derailment severity, both of which are subject to a variety of factors. Track class has been used as a proxy for track quality and as a parameter for estimating derailment rate because of its strong correlation with derailment rate. However, the relationship is complex. Some accident causes show a strong relationship, but others do not. On lower track classes track-related derailments are more likely to occur than derailments caused by equipment failures. Higher track classes are statistically associated with lower derailment rates for the majority of track-related and equipment-related accident causes. However, some equipment-related causes, notably hunting and truck structure defects, tend to have higher derailment rates and the corresponding higher risk on higher track classes. Although some equipment failures are common, such as bearing failures, they are likely to result in less severe derailments than track defects. Of particular interest are broken rails or welds, which are likely to cause more severe consequences. Speed was found to be a proxy to estimate the average number of cars derailed in an accident, but the specific severity and speed relationship varies among different accident causes. When all these factors are considered, there are

TABLE 8 Derailment Risk Reduction by Upgrade to Track Class, per Billion Car Miles

Cause Group	Class 2 to 3	Class 3 to 4	Class 4 to 5
Broken rails or welds (08T)	111.7	58.4	6.2
Track geometry (excl. wide gage) (04T)	45.3	13.9	2.4
All track causes	212.3	117.7	15.7
Bearing failure (10E)	-1.1	6.3	2.3
Broken wheels (12E)	-2.2	9.2	0.3
All equipment causes	59.8	41.6	7.4
All accident causes	396.3	200.8	26.7

interactions among accident causes that are differently affected by track class upgrade. Upgrading track class will generally reduce track-related derailment risk, but it increases the derailment risks pertaining to certain equipment-related causes. These and other factors need to be properly accounted for when evaluating the safety benefits and costs associated with infrastructure upgrade as a risk reduction strategy.

## ACKNOWLEDGMENTS

The first author was partially supported by a grant from the BNSF Railway. The authors are grateful to the following individuals who provided insights and thought-provoking discussion: Athaphon Kawprasert, Dingqing Li, Roy Brooks Wiley, and Huimin Wu. The authors also appreciate the helpful comments from several anonymous reviewers about the analysis.

## REFERENCES

1. FRA. *Track Safety Standards*. 49 CFR 213. U.S. Department of Transportation, 2003.
2. Nayak, P. R., D. B. Rosenfield, and J. H. Hagopian. *Event Probabilities and Impact Zones for Hazardous Materials Accidents on Railroads*. Report DOT/FRA/ORD-83/20. FRA, U.S. Department of Transportation, 1983.
3. Treichel, T. T., and C. P. L. Barkan. *Working Paper on Mainline Freight Train Accident Rates*. Research and Test Department, Association of American Railroads, Washington, D.C., 1993.
4. Anderson, R. T., and C. P. L. Barkan. Railroad Accident Rates for Use in Transportation Risk Analysis. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1863, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 88–98.
5. Saat, M. R., and C. P. L. Barkan. Tank Car Safety Design vs. Infrastructure Improvements in Reducing Hazardous Materials Transportation Risks. Presented at INFORMS Annual Meeting, Pittsburgh, Pa., 2006.
6. Lai, Y.-C., M. H. Dingler, and S.-Y. Lin. Optimization of the Railroad Network Track Class. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.
7. Liu, X., M. R. Saat, and C. P. L. Barkan. Benefit–Cost Analysis of Infrastructure Improvement for Derailment Prevention. *Proc., ASME-IEEE-ASCE-AREMA-TRB Joint Rail Conference*, University of Illinois at Urbana–Champaign, Urbana, 2010.
8. Kawprasert, A. *Quantitative Analysis of Options to Reduce Risk of Hazardous Materials Transportation by Railroad*. PhD dissertation. University of Illinois at Urbana–Champaign, Urbana, 2010.
9. *Special Study: Risk Concepts in Dangerous Goods Transportation Regulations*. NSTB-STB-71-1. National Transportation Safety Board, Washington, D.C., 1971.
10. Center for Chemical Process Safety. *Guidelines for Chemical Transportation Risk Analysis*. American Institute of Chemical Engineers, New York, 1995.
11. Anderson, R. T. *Quantitative Analysis of Factors Affecting Railroad Accident Probability and Severity*. MS thesis. University of Illinois at Urbana–Champaign, Urbana, 2005.
12. Saat, M. R., and C. P. L. Barkan. The Effect of Rerouting and Tank Car Safety Design on the Risk of Rail Transport of Hazardous Materials. *Proc., 7th World Congress on Railway Research*, Montreal, Quebec, Canada, 2006.
13. Kawprasert, A., and C. P. L. Barkan. Effects of Route Rationalization on Hazardous Materials Transportation Risk. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2043, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 65–72.
14. Glickman, T. S., and D. B. Rosenfield. Risks of Catastrophic Derailments Involving the Release of Hazardous Materials. *Management Science*, Vol. 30, No. 4, 1984, pp. 503–511.
15. Hauer, E. On the Estimation of the Expected Number of Accidents. *Accident Analysis and Prevention*, Vol. 18, No. 1, 1986, pp. 1–12.
16. Erkut, E., and A. Ingolfsson. Transport Risk Models for Hazardous Materials: Revisited. *Operations Research Letters*, Vol. 33, 2005, pp. 81–89.
17. Barkan, C. P. L., C. T. Dick, and R. Anderson. Railroad Derailment Factors Affecting Hazardous Materials Transportation Risk. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1825, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 64–74.
18. Anderson, R. T., and C. P. L. Barkan. Derailment Probability Analyses and Modeling of Mainline Freight Trains. *Proc., 8th International Heavy Haul Railway Conference*, International Heavy Haul Association, Rio de Janeiro, Brazil, 2005.
19. Schafer, D. H., II, and C. P. L. Barkan. Relationship Between Train Length and Accident Causes and Rates. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2043, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 73–82.
20. *Impact of Curvature on Track Geometry Safety Standards*. CSTT-RVC-TR-010. National Research Council Canada, Ottawa, Ontario, Canada, 2003.
21. *Risks Involved in Setting and Maintaining Speed Restrictions*. Arthur D. Little, Ltd., Cambridge, United Kingdom, 2004.
22. Lahrech, Y. *Development and Application of a Probabilistic Risk Assessment Model for Evaluating Advanced Train Control Technologies*. MS thesis. Massachusetts Institute of Technology, Cambridge, 1999.
23. John Volpe National Transportation Systems Center. Base Case Risk Assessment: Data Analysis and Tests. Office of Safety Analysis, FRA RSAC/PTC Working Group Risk 2 Team, 2003.
24. Dennis, S. M. Changes in Railroad Track Accident Rates. *Transportation Quarterly*, Vol. 56, No. 4, 2002, pp. 161–174.
25. Office of Safety Analysis, FRA. *Download Data on Demand*. [http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on\\_the\\_fly\\_download.aspx?itemno=3.03](http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx?itemno=3.03). Accessed June 6, 2010.
26. Office of Safety Analysis, FRA. *Train Accident Cause Codes*. <http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/downloads/appendixC-TrainaccidentCauseCodes.aspx?State=0>. Accessed June 6, 2010.
27. *Risk Assessment for the Transportation of Hazardous Materials by Rail, Supplementary Report: Railroad Accident Rate and Risk Reduction Option Effectiveness Analysis and Data*, 2nd revision. Arthur D. Little Inc., Cambridge, Mass., 1996.
28. Alali, A. *An Empirical Characterization of Commits in Software Repositories*. MS thesis. Kent State University, Ohio, 2008.
29. Steets, P. G., and Y. H. Tse. Conrail's Integrated Automated Wayside Inspection. *Proc., IEEE/ASME Joint Railroad Conference*, Piscataway, N.J., 1998, pp. 113–125.
30. Bladon, T. Predictive Condition Monitoring of Railway Rolling Stock. *Proc., Conference on Railway Engineering*, Darwin, Australia, 2004.
31. Barke, D., and W. K. Chiu. Structural Health Monitoring in the Railway Industry: A Review. *Structural Health Monitoring*, Vol. 4, No. 1, 2005, pp. 81–93.
32. Robeda, J., and S. Kalay. Technology Drives U.S. Train Inspections. *International Railway Journal*, Vol. 48, No. 5, 2008, pp. 47–50.
33. Lagnebäck, R. *Evaluation of Wayside Condition Monitoring Technologies for Condition-Based Maintenance of Railway Vehicles*. Licentiate thesis. Luleå University of Technology, Luleå, Sweden, 2007.
34. Robert, W., A. Aeppli, and P. Little. *Post-Audit of Wayside Detector Costs and Benefits*. Cambridge Systematics, Inc., Cambridge, Mass., 2009.
35. Schlake, B. W., C. P. L. Barkan, and J. R. Edwards. Train Delay and Economic Impact of In-Service Failures of Railroad Rolling Stock. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2261, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 124–133.
36. El-Sibaie, M., and Y.-J. Zhang. Objective Track Quality Indices. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1863, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 81–87.
37. Saccomanno, F. F., and S. El-Hage. Minimizing Derailments of Railcars Carrying Dangerous Commodities Through Effective Marshaling Strategies. In *Transportation Research Record 1245*, TRB, National Research Council, Washington, D.C., 1989, pp. 34–51.
38. Zarembski, A. M., and R. R. Resor. *Technical Monograph: Estimating Maintenance Cost for Mixed High-Speed Passenger and Freight Rail Corridors*. Report. FRA, U.S. Department of Transportation, 2004.



**TRANSPORTATION RESEARCH RECORD:  
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD**

## **Peer Review Process**

The *Transportation Research Record: Journal of the Transportation Research Board* publishes approximately 25% of the more than 3,900 papers that are peer reviewed each year. The mission of the Transportation Research Board (TRB) is to disseminate research results to the transportation community. The Record series contains applied and theoretical research results as well as papers on research implementation.

The TRB peer review process for the publication of papers allows a minimum of 30 days for initial review and 60 days for rereview, to ensure that only the highest-quality papers are published. At least three reviews must support a committee's recommendation for publication. The process also allows for scholarly discussion of any paper scheduled for publication, along with an author-prepared closure.

The basic elements of the rigorous peer review of papers submitted to TRB for publication are described below.

### **Paper Submittal: June 1–August 1**

Papers may be submitted to TRB at any time. However, most authors use the TRB web-based electronic submission process available between June 1 and August 1, for publication in the following year's Record series.

### **Initial Review: August 15–November 15**

TRB staff assigns each paper by technical content to a committee that administers the peer review. The committee chair assigns at least three knowledgeable reviewers to each paper. The initial review is completed by mid-September.

By October 1, committee chairs make a preliminary recommendation, placing each paper in one of the following categories:

1. Publish as submitted or with minor revisions,
2. Publish pending author changes and rereview, or
3. Reject for publication.

By late October, TRB communicates the results of the initial review to the corresponding author indicated on the paper submission form. Corresponding authors communicate the information to coauthors. Authors of papers in Category 2 (above) must submit a revised version addressing all reviewer comments and must send an e-mail to the chair or peer review coordinator, explaining how the concerns have been addressed.

### **Rereview: November 20–January 25**

The committee chair reviews revised papers in Category 1 (above) to ensure that the changes are made and sends the Category 2 revised papers to the initial reviewers for rereview. After rereview, the chairs make the final recommendation on papers in Categories 1 and 2. If the paper has been revised to the committee's satisfaction, the chair will recommend publication. The chair communicates the results of the rereview to the authors.

### **Discussions and Closures: February 1–May 15**

Discussions may be submitted for papers that will be published. TRB policy is to publish the paper, the discussion, and the author's closure in the same Record.

Many papers considered for publication in the *Transportation Research Record* are also considered for presentation at TRB meetings. Individuals interested in submitting a discussion of any paper presented at a TRB meeting must notify TRB no later than February 1. If the paper has been recommended for publication in the *Transportation Research Record*, the discussion must be submitted to TRB no later than April 15. A copy of this communication is sent to the author and the committee chair.

The committee chair reviews the discussion for appropriateness and asks the author to prepare a closure to be submitted to TRB by May 15. The committee chair reviews the closure for appropriateness. After the committee chair approves both discussion and closure, the paper, the discussion, and the closure are included for publication together in the same Record.

### **Final Manuscript Submittal: March 15**

In early February, TRB requests a final manuscript for publication—to be submitted by March 15—or informs the author that the paper has not been accepted for publication. All accepted papers are published by December 31.

### **Paper Awards: April to January**

The TRB Executive Committee has authorized annual awards sponsored by Groups in the Technical Activities Division for outstanding published papers:

- Charley V. Wootan Award (Policy and Organization Group);
- Pyke Johnson Award (Planning and Environment Group);
- K. B. Woods Award (Design and Construction Group);
- Patricia F. Waller Award (Safety and System Users Group);
- D. Grant Mickle Award (Operations and Preservation Group); and
- John C. Vance Award (Legal Resources Group).

Other Groups also may nominate published papers for any of the awards above. In addition, each Group may present a Fred Burggraf Award to authors 35 years of age or younger.

Peer reviewers are asked to identify papers worthy of award consideration. Each Group reviews all papers nominated for awards and makes a recommendation to TRB by September 1. TRB notifies winners of the awards, which are presented at the following TRB Annual Meeting.

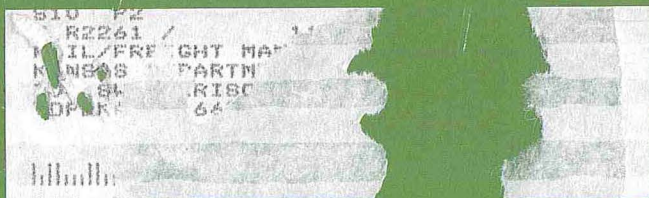
**Transportation Research Board**  
**www.TRB.org**

**TRANSPORTATION RESEARCH BOARD**

500 Fifth Street, NW  
Washington, DC 20001

ADDRESS SERVICE REQUESTED

NON-PROFIT ORG.  
U.S. POSTAGE  
PAID  
WASHINGTON, D.C.  
PERMIT NO. 8970



**THE NATIONAL ACADEMIES™**

*Advisers to the Nation on Science, Engineering, and Medicine*

The nation turns to the National Academies—National Academy of Sciences, National Academy of Engineering, Institute of Medicine, and National Research Council—for independent, objective advice on issues that affect people's lives worldwide.

[www.national-academies.org](http://www.national-academies.org)

ISBN 978-0-309-22296-9

90000



9 780309 222969